

Financial flexibility the key to Germany's scientific prosperity

Ten years after the fall of the Berlin Wall, and despite a traumatic transition, science in east Germany is flourishing. Nationwide obstacles to progress require increased federal leverage for their removal.

A glance across Central and Eastern Europe quickly reveals how well east German science has done since the Berlin Wall fell ten years ago. Without the generous subsidies from the German federal government after reunification and the tough imposition of Western standards of research, east German research institutes and universities would probably be in the same stagnant state as most of the cash-starved research institutes further east. Instead, many are in a position to compete with the best in the world.

But there was pain and also significant cost in the transition (see Briefing, page 635), including the loss of the exceptionally high quality of teaching in the east German university system. And not all of the many dismissals were fair. But now the need must be to provide healthy and, especially, flexible funding for science in the east as well as the west of the country, based on assessments of quality that can no longer be so predominantly focused on the east–west axis.

The high financial cost of the absorption of east German research has already begun a much-needed shake-up in the west German system, where scientists found themselves facing unfamiliar funding cuts in the 1990s, and under consequent pressure to submit to evaluations of their work and institutes, just as their eastern colleagues had to do.

The evaluation fever now raging through German science will benefit both east and west. Every publicly funded research institute and national research centre has undergone, or is going through, well-organized external evaluations based on widely accepted criteria. Even the major research organizations are being evaluated to check that their distribution of funds is as efficient as possible. The Wissenschaftsrat, Germany's science council, will integrate the results in a report on the entire research landscape next summer. Research minister Edelgard Buhllman must then decide what steps to take to improve the operation of the research system as a whole.

To bring this shake-up to its logical conclusion, the eye-glazing and overly complex arrangements of financing must be addressed. The current system, under which research organizations are jointly financed by federal and *Länder* (state) governments, is so inflexible that it is almost impossible to shift funds from one organization to another, or even within an organization. This is not least because the proportions of contributions of the partners vary enormously. For example, the federal government foots 90 per cent of the bills from the national research centres and the applied research institutes of the Fraunhofer Society, but only 50 per cent of those from the Max Planck Society and the Leibniz Society. Although an evaluation in the early 1990s recommended the break-up of one national research centre into smaller units that would be transferred to the Fraunhofer Society or the Leibniz Society, existing financing mechanisms have so far blocked the change.

The research ministry is keen to simplify the system, but cannot do so independently of the more general and even more complex system of tax redistribution (*Finanzausgleich*) between *Länder*, in which rich *Länder* must redistribute a proportion of their federal receipts to poorer *Länder*. The highly sensitive *Finanzausgleich* mechanisms are currently being discussed at government level, so this is exactly the time to broach the issue of a single financing mode for research organizations.

Most German science policy experts believe that a 50/50 split between federal and *Länder* governments across the board would be the right way forward. This would ensure an even balance of political influence over individual research institutes, and it fits well with traditional German consensus-seeking policies. However, given the need for tough decisions to be made quickly, it would be wiser to push for a split — say 60/40 — that would introduce a power balance in favour of the federal government while still leaving a reasonable level of influence for the *Länder*. ■

Caution: traditional knowledge

Principles of merit need to be spelt out in distinguishing valuable knowledge from myth.

What makes knowledge 'scientific', and thus distinguishable from other forms? This question has a relevance well beyond the philosophical, with the scientific nature of data having an increasing significance in issues ranging from the reproducibility of discoveries for which patents are being applied, to the argument that environmental regulations should be based on 'sound science'.

The question also overshadows the growing recognition that 'traditional knowledge', such as folk remedies for illnesses, deserves greater respect from modern science than it often receives. But such acceptance also requires due caution, and a rigorous assessment of more and less deserving forms of traditional knowledge.

That critically minded agenda may now be about to be established. Member organizations of what was previously known as the International Council of Scientific Unions (ICSU) have made clear their reservations about an open-ended endorsement of traditional

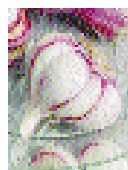
knowledge systems. At their general assembly in Cairo last month, they asked ICSU to carry out a "critical study" of the idea of traditional knowledge, concerned at the all-embracing interpretations that might be made of the concept, and particularly the support it might be taken to give to ideas such as astrology and creationism (see page 631).

The task should not be approached naively. Superficially attractive solutions to distinguishing between forms of knowledge, such as dismissing some knowledge systems (like astrology) on the grounds of non-reproducibility, could also rule out valid scientific ideas (such as natural evolution). But a clear statement of the practical principles on which scientific knowledge is based would be widely welcomed, especially if it also helps those battling with anti-scientific dogmas in both developed and developing countries. ICSU has yet to demonstrate its ability to live up to its new title, the International Council for Science. This is an opportunity for it to do so. ■





Astronomy
International array
project looks set to
sign up Japan
p627



Inquiry
Data in paper on the
medical benefits of
garlic questioned
p629



Knowledge war
Will 'traditional
knowledge' be used
to attack science?
p631



Drug doubts
UK questions cost-
effectiveness of flu
drug Relenza
p631

Medicine Nobel goes to pioneer of protein guidance mechanisms

London

This year's Nobel Prize for Physiology or Medicine has been awarded to Günter Blobel, of the Rockefeller University in New York, for his discovery that proteins are manufactured with in-built address labels that guide them to their destinations in the cell.

Cells contain compartments in which proteins perform distinct tasks. For example, DNA is transcribed in the nucleus, energy is generated in mitochondria, and the export of proteins out of the cell involves the endoplasmic reticulum.

The boundaries of these compartments are formed by lipid membranes that are largely impermeable to proteins. Most proteins, however, are made in the fluid surrounding the compartments, the cytoplasm. This raises the questions: How do these proteins know where to go? And how do they make their way through the lipid membrane?

Blobel found answers to both questions. Born in Germany in 1936, he took his MD at the University of Tübingen and his PhD at the University of Wisconsin in Madison. He joined the laboratory of George Palade — a Nobel laureate himself — at Rockefeller University as a postdoctoral fellow in 1967.

In 1971, with his then collaborator David Sabatini, Blobel suggested that secreted proteins contain intrinsic signals that deliver them to and across the endoplasmic reticulum membrane.

"These targeting signals are genetically encoded. They are built into the amino-acid sequences of the protein, and are an integral part of its structure. It's a brilliant and imaginative concept," says Gideon Dreyfuss, of the Howard Hughes Medical Institute at the University of Pennsylvania, Philadelphia. "It turned out to be correct, and spawned much of today's research on intracellular transport."

Blobel wrote in an article in *Current Contents* in 1985 that the "first experimental landmark" supporting his theory was the finding by César Milstein and his colleagues of a transient amino-acid sequence at the end of a protein that was secreted from the cell.

Based on this finding and their own evidence, Blobel and Bernard Dobberstein formulated in 1975 their 'signal hypothesis'.



Blobel: 'trailblazer' in molecular cell biology.

This predicted — in breathtaking detail — how newly formed proteins were targeted to the endoplasmic reticulum and entered it through a channel in the membrane. Sabatini, now at New York University, says: "One of my greatest satisfactions was to see some of those beautiful ideas substantiated by elegant and rigorous experiments carried out in Günter's laboratory."

William Wickner of Dartmouth Medical School in Hanover, New Hampshire, says: "Blobel predicted that there would be specific sorting receptors that recognized the signals. He went on to discover one such recognition system, the signal recognition particle and its membrane receptor, and to find that this system can coordinate protein synthesis and translocation. This work set the standard for studies in the fields of translocation and protein trafficking."

Since then, it has been shown that Blobel's signal hypothesis was correct and universal:

different organelles import proteins in a similar manner, and the process is essentially identical in yeast, plant and animal cells.

Today cell biologists "take the concept of intrinsic signals for granted", says Harris Bernstein of the US National Institutes of Health. "It's so largely engrained that people forget about where it came from." But, he adds, "it was revolutionary in its day".

The Nobel prize did not come as a surprise to Blobel's peers. Most consider it well deserved and long overdue (Blobel was awarded the Lasker Prize in 1993). "For 20 years Günter Blobel has been the trailblazer and pace setter in a major field in cellular and molecular biology," says a delighted Palade. Angus Lamond of the University of Dundee, Scotland, describes Blobel as one of the key figures who has transformed 'cell biology' into 'molecular cell biology'.

Palade adds that equally worthy of high recognition is Blobel's success in training an impressive group of young investigators. "Among them are the new leaders, or the promising future leaders of the field," Palade says. "This is the beginning of a dynasty." This group includes Dobberstein, who currently works at the University of Heidelberg, Larry Gerace of the Scripps Research Institute in La Jolla, California, and Vishwanath Lingappa and Peter Walter, both at the University of California in San Francisco.

Lingappa fondly recalls the night when Blobel returned to the lab after a smart party to finish his experiments — still wearing his tuxedo. "Such was his complete focus on the problem, the passion and his commitment to solving it."

Walter puts it slightly differently: "He has never shied away from controversy, and has aided the field by adhering to clear and testable hypotheses. Often he appeared dogmatic, yet he has always been [open to being convinced] by clean and unrefutable data. I couldn't have wished for a better mentor when I was a graduate student in his lab."

Blobel reportedly plans to donate part of his \$960,000 prize money towards the rebuilding of the Frauenkirche cathedral, which was destroyed in the bombing of Dresden in 1945.

Marie-Thérèse Heemels

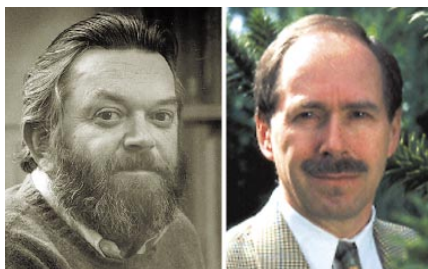
Dutch theoreticians win physics prize

London

This year's Nobel Prize for Physics has gone to Gerardus 't Hooft of the University of Utrecht and Martinus Veltman, emeritus professor at the same university, who showed how the complex mathematical structure of the theory of the electroweak interaction need not be an obstacle to predictive calculations.

They showed how to perform precise calculations of the physical properties of the new particles predicted by this and related theories, many of which have now received experimental confirmation. "The major advances in particle physics of the past 25 years depend on their work," says Christopher Llewellyn-Smith of University College London.

Initial progress with the theory of the electroweak interaction was hampered by the same problem that had plagued the development of quantum electrodynamics (QED) — how to cope mathematically with the short-lived particles and anti-particles that quantum theory predicts will surround



Veltman (left) and 't Hooft: Dutch double success.

a 'naked' central particle.

For QED, the problem was solved by the introduction of a 'renormalization' procedure: the experimentally determined values of mass and charge for the particle are used as 'renormalized' input parameters.

But the electroweak theory does not possess the symmetry simplicity of QED. The particles mediating the electromagnetic interaction are the massless, non-interacting photons; for the electroweak interaction, however, they are massive and interacting. The mathematical difficulties arising from

these additional properties were such that many theorists at the time doubted the physical viability of the electroweak theory.

Veltman, optimistic that the calculational difficulties could be overcome, developed the computational tools and showed that many of the more problematic aspects of the electroweak theory effectively cancelled each other out.

According to Llewellyn-Smith, "Veltman was the driving force keeping interest in such theories alive." 't Hooft, Veltman's student at the time, had the insight that if the masses of particles such as W and Z were a natural consequence of the theory, rather than input parameters, then the theory became self-consistent and tractable.

These combined developments were key, and the theory of the electroweak interaction became, by the early 1970s, a practical — and very successful — theoretical tool for performing precise calculations. Calculations of physical quantities associated with the W and Z particles and the top quark show agreement with recent measurements.

Karl Ziemelis

Publishers agree on a 'seamless web'

London

The publishers of the world's leading scientific journals have agreed to adopt common procedures that will allow a 'seamless web' linking references in the articles they publish to the source papers in their respective publications through the World-Wide Web.

The agreement was reached in Frankfurt on Tuesday by 300 industry executives attending a meeting of the International Association of Science, Technology and Medical Publishers. It is the result of discussions and trials that have been taking place over the past two years.

The purpose of what is described as "the world's largest research web" will be to ensure that any paper cited in the electronic versions of the journals covered — including the *Nature* family of journals — will be directly accessible via the web.

It is expected that eventually between five and ten million scientific articles and their references will be interconnected in this way. A final plan is to be discussed at a meeting in London in December.

"This simple act begins to deliver on the full promise of the web," says Robert Campbell, chair of the association. "We are working together to serve our primary constituencies: the authors and readers who comprise the world community of scientists."

David Dickson

Split-second chemistry is rewarded

London

The Nobel Prize for Chemistry has been awarded to Ahmed Zewail of the California Institute of Technology for his role in developing techniques of ultrafast spectroscopy for studying molecular change in real time.

The fundamental time-scale of chemical reactions is set by the speed at which the reactant molecules pass through the transition state. This involves atomic motions which typically occupy just a few tens of femtoseconds (10^{-15} s).

Short pulses of laser light are used to capture this change. A series of femtosecond pulses provides a strobed sequence of the molecules' transformation. Each pulse records a slice through the absorption spectrum, modified by the change in the molecule.

This process requires, however, that one knows a 'zero' time from when the change began. This involves synchronizing two pulses with femtosecond precision. The initial 'pump' pulse provides the stimulus for a photochemical process — for example, boosting a collection of molecules into an excited state. Subsequent pulses then interrogate the molecules as they pass through the transition state to the products.

Zewail has pioneered the implementation of this scheme. His early experiments focused on the simplest case: unimolecular



Zewail: honoured in his native Egypt.

photodissociation, whereby a molecule absorbs a photon and fragments into two. His later work encompasses more complex reactions, such as the dissociation of $C_2I_2F_4$ into three fragments, bimolecular processes such as the reaction of carbon dioxide with a hydrogen radical, and

the influence of solvent cages on reaction dynamics in solution.

John Simons of the University of Oxford says that Zewail has been "marvellously effective in demonstrating the potential of ultrafast laser techniques — of showing quantum behaviour happening before your eyes". These methods are now being applied beyond elementary test cases, for example to elucidate some of the rapid electron transfer processes in photosynthesis.

It seems inevitable that ultrafast change in biological systems will receive increasing attention. Another goal is the use of pulsed laser techniques for ultrafast diffraction, so that precise changes in atomic coordinates can be tracked during chemical change.

Zewail has been a favourite for the award for several years.

Philip Ball

US State Department hires consultant to boost role of science

Washington

The US Department of State announced last week that it will employ Jack Gibbons, former science adviser to President Bill Clinton, as a consultant to help guide its consideration of a critical report from the National Research Council (NRC) on its understanding and use of science.

The report, which reflects pressure from both the scientific community and Congress, calls for extensive changes in training and organization to boost the role of science and technology in US foreign policy. The department's move is seen as a commitment to achieve this.

Madeleine Albright, the secretary of state, also promised to meet with Bruce Alberts, the president of the National Academy of Sciences, to discuss the report, as soon as her schedule permits. Alberts first asked for such a meeting a year ago.

The State Department has promised many times to address accusations that it ignores science, but changes have been slow in coming. For example, senior officials have pledged twice in the past two years that a senior scientific adviser would be appointed, but no such appointment has been made.

The drive for reform seems to have gained new impetus after the Senate passed a foreign-relations bill requiring the department to appoint a science adviser and to report back to the Congress on how it plans to implement the NRC's recommendations. The bill is expected shortly to become law.

The NRC says the State Department should train all foreign-service officers and diplomats to give them "basic competence in scientific, technological and health matters", and that it should appoint 25 scientifically qualified senior diplomats at its foreign embassies.

But Glenn Schweitzer, the director of the NRC study, says its most important recommendation is that the secretary of state should establish a science and technology policy, together with an "action agenda" to implement it.

"We're pleased that the committee has done such a thorough job," says Ken Brill, acting assistant secretary for oceans, environment and science at the State Department. Brill says the department is facing "a new set of issues, many of them science-based" and that it plans to respond to the challenge. Gibbons, he says, will help the department come up with a science adviser's position "that will be effective".

Colin Macilwain

Varmus announces decision to quit NIH for cancer centre

Washington

Harold Varmus, director of the US National Institutes of Health (NIH), announced last week that he will resign at the end of the year to become president of the Memorial Sloan-Kettering Cancer Center in New York.

Varmus's six years in office have seen a remarkable expansion in NIH's budget — from just over \$10 billion to almost \$16 billion — and something of a renaissance in the scientific reputation of the institute's intramural programmes.

He claims that "the whole enterprise is a happier and a stronger place" than when he arrived, partly because of the budget but also because of the people he has recruited.

Varmus's letter of resignation to President Bill Clinton called on the president "to consider other active medical scientists" in the search for a successor.

Varmus, who won the Nobel prize for physiology and medicine in 1989 with Michael Bishop for his work on retroviral oncogenes, says his top priority in New York will be "bringing basic science together with the clinical researchers". He intends to strengthen some areas of research, including cancer genetics, and to collaborate with other New York medical centres.

During his tenure at NIH, Varmus developed a knack for preventing difficult issues, such as embryo research, from disrupting the agency's ties with the Congress. He also maintained excellent relations with Capitol Hill after Republicans assumed control of both houses of Congress in 1995.

Ruth Kirschstein, NIH deputy director and a long-time NIH administrator, will take over as acting director while a search committee seeks a nominee to replace Varmus.

With the presidential elections due next November, it may prove difficult for Clinton and Donna Shalala, the health secretary and Varmus's immediate boss, to find a suitable replacement. Any appointee would have to be confirmed by the Republican-controlled Senate, and be prepared to work for a new health secretary in a new administration, Democrat or Republican.

This may make the position unattractive to an outsider in the Varmus mould, and early speculation has focused on top scientific administrators already at NIH who have good political connections with both parties. These include Tony Fauci, director of the National Institute of Allergy and Infectious Diseases, and Richard Klausner, head of the National Cancer Institute.

Colin Macilwain

Japan plans to join array project

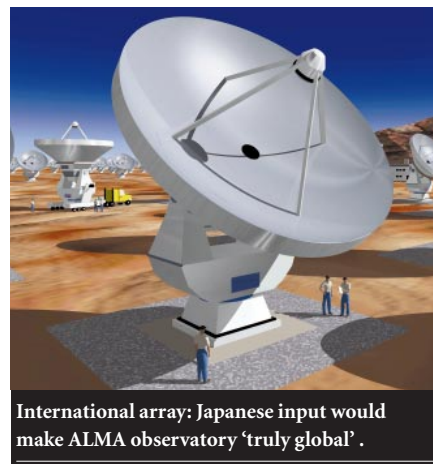
Tokyo

Japan is planning to become the third partner of the Atacama Large Millimeter Array (ALMA), an international project currently led by US and European astronomers to build a large millimetre-wave observatory at a mountain site 5,000 metres high in Chile's Atacama Desert.

The collaboration would be based on a joint framework discussed during a meeting in Washington last week, at which ALMA partners and Japanese astronomers explored scientific aspects of the project.

Riccardo Giacconi, the president of Associated Universities Inc., which operates the US National Radio Astronomy Observatory (NRAO) — responsible on behalf of the National Science Foundation for the US end of ALMA — says that Japan's involvement would create "the world's first truly global astronomical collaboration".

At a meeting in Tokyo two weeks ago, the three partners discussed scientific and technological aspects of the planned collaboration and agreed that merging ALMA with Japan's planned Large



International array: Japanese input would make ALMA observatory 'truly global'.

Millimetre and Submillimetre Array (LMSA) would lead to a better, more sophisticated facility.

An agreement was signed in Washington in June by US and European representatives to continue collaboration on the first phase of the telescope, which will image the Universe at millimetre wavelengths, between the radio and infrared spectral regions. It is

► hoped that full approval for the project will be obtained by 2001, with limited operations starting in 2005, and that the array will be fully operational by 2009.

Keiichi Kodaira, director-general of Japan's National Astronomical Observatory (NAO), which oversees LMSA, argues that Japan's involvement will bring substantial benefits to the project. "Since LMSA is already in the advanced stages of its design and development phase, there is much that Japan could contribute to the development of ALMA," says Kodaira.

But he says that Japan would not be able to join ALMA as an official partner until it can secure the necessary funding from the Ministry of Education, Science, Sports and Culture (Monbusho), which oversees NAO.

"One of the problems is the fact that we have a very different funding system from that of US and European research institutes," says Kodaira. "Since the budget request for next year has already closed, we hope to secure relevant funding after the merger of Monbusho and the Science and Technology Agency (STA) in 2001."

Until then, LMSA and ALMA are expected lead a separate existence, although the three sides have agreed on Japan's early involvement in the project at a scientific and technological level. Negotiations are currently taking place to add Canada to the US team, and Spain to Europe's participation.

Kodaira says that Japan will participate in the design and development phase of ALMA, including the investigation of the site.

Last week's meeting included discussions on the design differences between LMSA, which was planning to build 50 ten-metre antenna dishes, and ALMA, which plans to build 64 telescopes with 12-metre antennas. The purpose of the meeting was "to plan how to maximize ALMA's scientific potential", says Paul Vanden Bout, director of NRAO.

ALMA, whose members also include the European Southern Observatory, France's Centre National de la Recherche Scientifique and the United Kingdom Particle Physics and Astronomy Research Council, is to study the structure of the early universe, the formation of the stars and the evolution of galaxies.

Kodaira hopes that ALMA will change the general perception in Japan that astronomy has no practical applications. He says that research into millimetre-wave interferometry could lead to new technologies in telecommunications, information technology and semiconductors, all priority areas in next year's budget. **Asako Saegusa & Natasha Loder**

Aim for better business sense to bolster Russian science

Moscow

The Russian Academy of Science (RAS) has set up a new body, known as the Above-Budget Foundation (ABF), to manage its property and channel the income it receives to supporting scientists and their work.

According to Gennady Mesyats, the academy's vice-president responsible for ABF activities, the new body has already generated more than \$1 million in income, even though it started operating only a few months ago.

"Next year we hope to get \$3 million, and in the coming years ABF plans to operate with tens of millions of dollars," says Mesyats. "We hope that this will enable our libraries to subscribe to all major journals, and that we can finance expeditions, and solve our most acute financial problems."

Mesyats told a meeting of the academy's praesidium in Moscow last week that, for the first time in recent years, the government has this year transferred all the money it had allotted to science both on time and in full. But the praesidium emphasized that the total amount being made available is still insufficient to meet the needs of Russian science.

External income is therefore very important, Mesyats emphasized, particularly given that the budget money can only be used for specified purposes, such as salaries, and cannot be redistributed. The academy owns considerable property, which until recently has been inefficiently used. It has 38 buildings in Moscow alone, and many stores, hotels and other real estate around the country.

The academy also receives income from its scientific ships, and substantial revenue is raised through its international academic association 'Science', which manufactures



and distributes copies of scientific exhibits.

The academy also makes money from publishing. More than 80 scientific journals are published in English by a recently founded publishing company, Nauka, which is expected to raise \$1 million next year, partly because publication of the two leading Russian physics journals was transferred to it from the American Institute of Physics.

The ways in which these funds are to be spent are discussed at weekly meetings of the RAS praesidium, chaired by Yuri Osipov, the academy's president. This year's priorities have included buying apartments for young scientists, providing petrol supplies for mountain observatories and supporting scientists' widows.

"Finding external funds is a new activity for us, even though in other countries it is quite normal; up to 30 per cent of the income of US universities comes from using their property, such as stadiums, parking places and living quarters," says Mesyats. **Carl Levitin**

Anger at Israeli sex crimes DNA bank

Jerusalem

Israeli police have been accused by the Association for Civil Rights in Israel of illegally taking DNA samples from suspected sex offenders in their search for a serial rapist.

The police have taken saliva samples from previous suspects in other sex cases, even though they are not believed to have any connection with the serial rapist, who has attacked women in the Tel Aviv region during the past year. The samples are being used to produce DNA profiles for a police data bank.

But Dan Yakir, the counsel for the civil rights association, has written to Israel's attorney-general, Elyakim Rubinstein, claiming that this activity is "manifestly il-

legal" and "seriously violates the fundamental rights to freedom, privacy, and respect". He adds: "The police have no authority to summon for investigation a person against whom there is no specific suspicion."

Yakir also argues that the establishment of a DNA data bank on sex offenders is illegal without legislation on this. But a spokeswoman for the Ministry of Internal Security says DNA samples are taken only with subjects' written consent. She acknowledges that the samples are being taken not only to find the serial rapist, but also to "advance the investigations of other rapes". But she argues that the samples benefit those who give them by clearing them of the crimes. **Haim Watzman**

'Faked' rock-dating data charge dismissed...

San Diego

Arizona State University (ASU) has cleared geography professor Ronald Dorn of fabricating rock-dating data, prompting the National Science Foundation (NSF) to close its investigation of allegations of scientific misconduct in Dorn's federally funded studies.

Milton Glick, the university provost, issued a brief statement last week stating that Dorn, who is widely known for dating geological and archaeological sites, "did not commit scientific misconduct in collecting and preparing rock varnish samples for carbon dating".

The ASU investigation found that "the evidence did not support allegations that Dorn added coal and charcoal to rock varnish samples to hit predetermined 'target dates'". But the controversy over Dorn's dating technique continues, as he is suing a group of researchers who published an article in *Science* in June 1998 detailing how they found coal and charcoal-like material in 80 of Dorn's rock samples (see *Nature* **401**, 419; 1999).

Dorn's technique involves scraping the varnish off a rock surface, preparing the material and sending it to a radiocarbon laboratory for analysis to try to determine a date for the geological formation or adjacent archaeological discovery.

The authors of the *Science* article wrote that Dorn's varnish-dating technique was meaningless, given the presence of the materials, as any results would be a ratio of the two materials with differing ages. In 1996, some authors raised questions about Dorn's results with the NSF, which funded much of his studies.

In court and state records, Dorn alleges that the *Science* authors intentionally, "with evil motives and malice," misrepresented and manipulated data and contaminated samples to accuse him of professional misconduct. None of the authors of the *Science* article would comment, but they are reported to be surprised at the university's verdict.

Arizona administrators say the disagreement between Dorn and his critics is scientific, not legal. "The issue is resolvable and should be resolved in the normal arena of science, involving assessments, measurements and debate," says Eugene Levy, an astrophysicist and dean of the university's College of Science.

Neither Dorn nor his attorney in Phoenix have been available for comment in recent weeks, and ASU officials have declined to release any documents about the investigation or be interviewed.

Earlier this year, however, Dorn and two co-authors described the dispute (*American Indian Rock Art* **25**, 1-31; 1999) as the "sordid

side of science", adding that Dorn was the victim of a "scientific witch hunt". In the same article, it was noted that Dorn verbally began retracting his varnish-dating method in 1996, as the investigation was beginning.

According to NSF spokeswoman Mary Hanson, the foundation's Office of Inspector General closed its investigation late last month after reviewing the ASU investigation, which showed that "the preponderance of the evidence didn't support the allegations". The inspectors felt that "any remain-

ing issues would best be resolved in the scientific community", says Hanson.

Howard Schachman, a professor of biochemistry at the University of California, Berkeley, and a member of the American Association for the Advancement of Science's committee on scientific freedom and responsibility, says that, although he is not familiar with the details of the Dorn case, court action over such a dispute could "have a very serious chilling effect" on the publishing and debating of results.

Rex Dalton

... and German garlic study under scrutiny

Munich

A German university last week began a formal investigation into allegations that a prominent researcher was involved in the fabrication of data in a paper claiming an antisclerotic effect for a garlic preparation.

The paper was subsequently published in a journal of which one of its co-authors was acting as guest editor. But the main author is strongly denying any suggestions of impropriety.

The Humboldt University in Berlin is investigating allegations against Holger Kiesewetter, the director of the Institute of Transfusion Medicine, part of its medical school, and joint author of a paper published in the May issue of the medical journal *Atherosclerosis* (**144**, 237-249; 1999).

The paper is based on a clinical trial carried out between 1991 and 1996, and concludes that taking high-dose garlic pills "can reduce the increase in arteriosclerotic plaque volume by six to 18 per cent, or even effect a regression, within four years".

But an article in the *Süddeutsche Zeitung* has questioned the authenticity of two figures in the paper. The figures are described as "very precise" ultrasound images of the common carotid artery of a patient before and after the 48-month therapy, and are used as proof of the claimed reduction in plaque volume.



A bad smell? Medical benefits of garlic have been questioned.

Doubts have been triggered by the similarity of the pictures' details and angles of vision. According to experts, if the pictures had been taken at different times, such similarities would have been a remarkable coincidence.

"Although the pictures are apparently not identical or retouched, they are very likely to date from the same series of shots," says Ulf Rapp, a professor of biology and immunology at the University of Würzburg.

"We are taking the case very seriously," says Cornelius Frömmel, dean of research at the Humboldt University. He says he plans to ask Kiesewetter for the original prints of the pictures.

Kiesewetter, the senior author of the paper, is also accused of ignoring the high drop-out rate of patients unable to bear the smell of the garlic preparation. "The possible bias caused by this was not taken into account in the statistical analysis," says Jürgen Windeler, a biostatistician at the University of Heidelberg.

But Kiesewetter hotly denies the allegations, claiming that the figures are authentic and that the statistical analysis is "correct". He says he is prepared to hand over all his files to investigators, and is confident that the enquiry will clear his name.

Kiesewetter's paper was chosen for publication by a guest editor of the German Society for Atherosclerosis, Günter Siegel, a professor of physiology at the Free University of Berlin. Siegel, then president of the society, is a co-author of the paper.

Siegel did the statistical analysis, and he agrees that the drop-out rate should have been studied more thoroughly. "But it would certainly not have significantly altered the results, nor added new aspects to the study," he says.

The study was commissioned and financed by the Berlin-based pharmaceutical company Lichtwer Pharma, the producers of Kwai, Germany's leading garlic preparation. The company has been using the results in its marketing campaign, trumpeting the preparation's 'clinical relevance' in advertisements.

At a Lichtwer Pharma press conference in May, Kiesewetter recommended the use of Kwai to anyone over 40. A spokesman for the company says it is considering stopping the campaign.

Quirin Schiermeier

CORBIS/MICHELLE GARRETT

Berkeley puts \$500m into multidisciplinary approach to disease

San Diego

The University of California at Berkeley last week announced a \$500 million Health Sciences Initiative, which will bring together biologists, physicists, mathematicians and computer scientists to address human disease.

About \$300 million will be used to construct new laboratory buildings and \$200 million will be spent on researchers and new equipment.

Berkeley, which lacks a medical school is seeking to create an interdisciplinary initiative that produces both biological discoveries and human products, from artificial tissues to drugs.

Robert Tjian, a professor of molecular and cell biology, Howard Hughes Medical Institute researcher and co-leader of the initiative, says that many future breakthroughs in biology will occur "at the convergence of these diverse fields".

Thomas Budinger, chair of the new bioengineering department — a combined effort involving the Berkeley campus and the University of California at San Francisco — says it will "bring engineering to biology and biologists to the methods and resources of engineering".

A Whitaker Foundation grant is expected to fund an expansion in bioengineering students, say Berkeley officials. The initiative also includes the new Helen Wills Neuroscience Institute, named after the former tennis star, whose estate donated \$10 million.

Rex Dalton

Joint institute set to boost Austria's genome research

Munich

The Austrian Academy of Sciences and the pharmaceutical company Boehringer Ingelheim have announced plans to set up a high-profile Institute of Molecular and Cellular Bioinformatics (IMBA) in Vienna.

The new institute will work closely with Boehringer Ingelheim's Institute of Molecular Pathology (IMP) in Vienna, sharing its scientific advisory board and service facilities. The two institutes will form a new Genome Research Centre aimed at combining research in molecular biology and its potential medical applications.

The IMBA will be owned and financed by the Austrian Academy of Sciences, with money from the Austrian ministry for science and transport, and the city of Vienna will pay for the construction of a new building. Annual core funding will be about 100 million schillings (US\$8 million) a year.

Boehringer Ingelheim will have no direct control over the IMBA's research strategy, but will give it access to the IMP's scientific infrastructure, including sequencing facilities and electron microscopes.

All income from joint scientific research will be shared between the institutes in accordance with their input. Boehringer Ingelheim will have an option to buy the IMBA's share of an invention, or to license it.

The IMBA will house 60–80 scientists in a new building adjacent to the IMP. The first groups will start work in temporary accommodation within a year, and recruitment of scientific staff will start immediately.

Kim Nasmyth, the IMP's scientific director, will keep his position at the IMP, but will also be responsible for establishing the IMBA's scientific programmes during the recruitment phase, and for coordinating the direction of research at the two institutes.

He is keen to create innovative career structures at the IMBA, which, like the IMP, will have no tenured positions. There will therefore be "huge and continuous opportunities for young scientists", says Nasmyth.

Research at the IMBA will focus on human molecular and cellular biology and, in particular, on the medical implications of research results from the IMP, which has been successful in basic research on model organisms aimed at understanding the origins of diseases such as cancer.

Research at the IMBA may include the development of methods for analysing normal and diseased human tissue, and identifying human polymorphisms that contribute to multigene diseases. Its bioinformaticists will work on transforming raw sequence data into knowledge for medical use.

Bernd Wetzel, head of international research and development at Boehringer Ingelheim, welcomes the decision by the Austrian Academy of Sciences to collaborate so closely with a privately funded institute. "They have recognized the mood of the times," he says.

Werner Welzig, president of the academy, says the cooperation with Boehringer Ingelheim marks a "new departure" for biological science in Austria.

Quirin Schiermeier

Institut Pasteur names Kourilsky as new director

Paris

France's leading medical research institute, the Institut Pasteur, last week named Philippe Kourilsky as its new director. Kourilsky, a specialist in molecular immunology who has been with the institute for 27 years, will replace Maxine Schwartz, whose second six-year term ends in January.

Under Schwartz, the institute's budget increased significantly to FF998 million (US\$160 million) in 1998, and it became heavily involved in fields such as the origins and treatment of AIDS.

Scientists at the Pasteur say Kourilsky's largest challenge will be to maintain its prestige as a competitive institution, while staying true to its nineteenth-century mission as a research institution focusing on



Kourilsky: must maintain prestige.

infectious diseases and public health.

"His major mandate is to make sure that Pasteur remains a prima donna in microbiological research," says Simon Wain-Hobson, director of a molecular retrovirology laboratory the institute. "We have to modernize and at the same time maintain our history."

Pasteur researchers say that Kourilsky is well placed to do this. He has been associated with the institute since 1972, but has also held posts at INSERM, France's

national biomedical research agency, and in industry, at the Lyons-based company Laboratoires Mérieux. He is currently a professor of molecular immunology at the Collège de France and a research director at the Pasteur.

Kourilsky is "both an insider — a true 'pasteurien' — and an outsider", says Jean-Louis Virelizier, director of a viral immunology lab at the institute. "He has all the potential to be a good director," says Wain-Hobson. "I think he is politically savvy, bright, dynamic. The real test will be when he gets to work."

Major tasks facing the director, in addition to coordinating Pasteur's 22 institutions scattered throughout the world, include defining its research strategy in areas such as genomics.

Heather McCabe

ICSU seeks to classify 'traditional knowledge'

London

The International Council for Science (ICSU), which brings together the main professional bodies representing individual scientific disciplines, has been asked to carry out a study of the concept of 'traditional knowledge'.

Some members are concerned that a commitment to promote such knowledge, endorsed by participants to the World Conference on Science in Budapest in June, could be used to support the argument for a range of 'anti-science' ideas, ranging from astrology to creationism.

The endorsement of traditional knowledge comes in two documents approved at the end of the Budapest conference, which was organized jointly by ICSU and the United Nations Educational, Scientific and Cultural Organization (Unesco).

A paragraph in the first document, *Declaration on Science*, describes traditional and local knowledge systems as "dynamic expressions of perceiving and understanding the world [that] can make, and historically have made, a valuable contribution to science and technology". The second document, *Science Agenda*, calls on governments to "formulate national policies that allow a wider use of the applications of traditional forms of learning and knowledge". (See <http://helix.nature.com/wcs> for copies of



Roots of knowledge? A medicine man plies his trade at a market in Malawi.

both documents.) The inclusion of these statements followed pressure from several countries — India in particular — that greater recognition should be given to, for example, the potential medical value of traditional herbal treatments (see *Nature* 397, 376; 1999).

It reflects feeling in many such countries that the importance of this knowledge is being undermined by the teaching of Western

science and that, where this knowledge turns out to have commercial value, those who developed it are not being compensated.

But when the two documents were discussed in Cairo two weeks ago by the general assembly of ICSU, there was also concern that 'traditional knowledge' was insufficiently defined.

"This arose because of the difficulty of interpreting this phrase, and a feeling that part of the debate over such knowledge is directly linked to various 'anti-science' lobbies," says Brian Heap, foreign secretary of Britain's Royal Society. "We felt that there was a need for clarification."

Representatives of the International Astronomical Union are said to have emphasized the need to distinguish between astronomy and astrology. US participants expressed similar concern that the documents might be used to support creationist ideas.

Heap and others acknowledge that there are areas where 'folk remedies' have a sound scientific basis. Indeed, he points out that academies in India and China support study of the links between traditional knowledge and modern science.

The assembly's endorsement of the documents expressed reservations about the passages referring to traditional knowledge. It has asked the executive board of ICSU "to set up a critical study of this issue". **David Dickson**

UK government not convinced by claims for flu drug

London

The UK government agreed last week that physicians should be discouraged from prescribing a novel anti-viral drug, Relenza, on the grounds that it is not convinced of the drug's cost-effectiveness.

The government's announcement endorsed advice to physicians from the newly created National Institute of Clinical Excellence (NICE), set up to provide guidance on whether new therapies should be introduced into the health service.

But Glaxo Wellcome — maker of Relenza — and two other leading pharmaceutical companies want the government to reverse its decision and overrule NICE. They argue that the need to obtain the institute's approval is anti-innovative and an unjustifiable additional hurdle for new drugs to overcome.

Relenza prevents the spread of the virus to uninfected cells in the respiratory tract. It is the first in a class of compounds known as neuraminidase inhibitors, and the first anti-viral drug to be effective against all known strains of influenza A and B.

For influenza to spread, new virus particles must break out from infected cells. The enzyme neuraminidase breaks the bond holding the virus to the infected cell. Relenza inhibits neuraminidase activity, so retaining the bonds between sialic acid and cell surface proteins, and trapping the virus in sugars on the outside of the cell.

The influenza virus is highly mutable, and comes in many different strains. This has made it difficult to locate a stable target to prevent viral replication. After three decades, researchers discovered that the role of neuraminidase is constant. Once this had been identified, Relenza was designed using computational chemistry.

Persuading the government of the drug's value has proved a different matter. Glaxo Wellcome says the drug reduces the duration of flu, the severity of the symptoms and the risk of secondary infections. But NICE says that, if used within 48 hours of the onset of symptoms, Relenza only reduces the duration of the illness by 24 hours — not the two to three days claimed by the company.

Crucially, because of the design of the clinical trials used by the company, NICE concluded there was "insufficient information to judge the extent to which the frequency of serious secondary complications in high-risk groups ... might be reduced". These groups are currently targeted by National Health Service (NHS) vaccination programmes. And, NICE says, it is difficult to see how most flu patients could be diagnosed within 48 hours within the NHS. Although some clinical work has shown that Relenza can limit the spread of flu between individuals in closed settings, it is not currently licensed for prevention.

The drug may have been the victim of unfortunate timing. With NICE needing to gain the confidence of the health professions, the public and parliament as early as possible, it would have been difficult for the government to reject its first advice.

NICE's rapid review of Relenza will not prejudice the full appraisal procedure that Relenza and a similar drug produced by the company Hoffman-La Roche are to undergo next year. **Natasha Loder**

US space and science agencies escape threatened cash cuts

Washington The US space agency NASA and National Science Foundation (NSF) had most of their budget requests for science restored last week, as appropriators in the Senate and House of Representatives met to reconcile their spending bills for fiscal year 2000. The deal virtually assures that the two agencies will avoid the deep cuts that were threatened just a few weeks ago.

The NSF will receive \$3.9 billion, seven per cent more than last year, and \$21 million below the request.

The NSF's research and 'related activities' account gets \$2.9 billion, including \$105 million for information technology research and the full \$36 million request for a teraflop computing system.

The House-Senate 'conference' restored \$75 million of the \$120 million the Senate had trimmed from NASA's \$2.1 billion space science and technology account. But congressionally mandated spending on 'earmarked' projects will still squeeze existing science missions, including the Discovery planetary programme.

Vice-President Al Gore's proposed Earth-viewing Triana satellite is put off until the

year 2001, and will need a stamp of approval from the National Academy of Sciences before work can proceed.

Romanian academy wants to oust US embassy

Moscow The Romanian Academy of Sciences is demanding the return of property that it once owned, including a building in the capital, Bucharest, that is now occupied by the US embassy.

The property was taken over by the Communist government in 1948; in accordance with a restitutions bill passed by the lower house of the Romanian parliament, it should be returned to the former owner. The upper chamber of the parliament has not yet discussed the law.

Talent survey seeks France's brightest

Paris In a bid to "modernize the structures of French research and find how to spot young talent", Claude Allègre, France's science minister, has decided to go beyond "classical citation indices" and ask a select group of scientists worldwide who they reckon to be the best French scientists.

Members of science academies and winners of major science prizes have been

asked to list first-class French scholars and groups in their own and other fields; whether any of these have made a major contribution to science; and to identify promising young scholars. Answers are being collected and analysed by Vincent Courtillot, Allègre's director-general of research.

Hackers wreak havoc at Imperial College, London

London Hackers have kept 100 staff and students at Imperial College of Science and Technology, London, locked out of their computers since last Tuesday (5 October). The attacks came from what a member of staff described as "very experienced, professional hackers".

The working hours of one of the hackers — called Military — suggest he or she is based in the United States. Military is rumoured to have visited Stanford and the Naval Research Laboratories in the past week. Hackers were known to have been showing an interest in the cluster of powerful Sun workstations at Imperial in recent weeks.

Australia's science comes under review

Sydney Australia's Minister for Industry, Science and Resources, Nick Minchin, has ordered a wide-ranging review of national

science capability from the country's chief scientist, Robin Batterham. The review will measure the needs of public and private sectors. It will look at what characteristics the country's science base needs if it is to support the development of leading-edge industry.

Minchin says the review will examine "what contribution science can and should make to economic development and wealth creation," and expects it to influence an 'Innovation Summit' next February. Some believe the review will endorse moves to transfer responsibility for university research from the education ministry to Minchin's department.

Lawrence Livermore in the clear on radar technology

San Diego An outside task force last week cleared the Lawrence Livermore National Laboratory in California of breaking regulatory and statutory requirements in patenting and licensing a micro-impulse radar technology (see *Nature* 400, 6; 1999).

But the panel, examining a long-running dispute over the valuable technology, recommended new policies to address shortcomings in the laboratory's business and communication procedures. The task force was appointed last spring after complaints by a small Alabama firm, Time

Domain Corp., and Democratic members of the House of Representatives Committee on Science, who questioned the laboratory's handling of the technology.

\$2m boost for work on xenotransplants

London Xenotransplantation research at the company PPL Therapeutics has been awarded \$2 million from the US Advanced Technology Program to support the cloning of transgenic pigs with organs likely to be accepted by human recipients. The National Institute of Standards and Technology in the United States is funding the programme, which aims to reduce the risk of rapid rejection of transplanted pig organs.

The rejection of non-human organs means xenotransplantation cannot yet address the shortage of human organs for transplants. A recipient's immune system can often be triggered into rejecting a graft by the product of a single gene in donor cells, and PPL is trying to knock out the gene causing such hyperacute rejection.

Catalonia plans to put its academic know-how online

Barcelona The vice-chancellors of eight public and private Catalan universities have

agreed on a regional project for the electronic management and certification of their academic processes and the digitization of their scientific and teaching assets. The initiative, 'Digital University in Catalonia 1999–2003', has been put forward by the Information Society Commission of the Catalanian government, with a Ptas111 million (US\$740,000) budget. It aims for complete digitization within the four year period.

Universities will be able to share information needed for undergraduate teaching, and to set up an Internet platform to facilitate access to university publishers. The move, approved by the Universities and Research Commission, will include digitization of university libraries, including the provision of a single server to act as a repository of doctoral theses. International Internet connections will also be improved to reach a minimal flow of 34 Mbps.

Correction

An article on restructuring within the Natural Environment Research Council's Centre for Coastal and Marine Sciences incorrectly stated that redundancies have been identified in the areas of numerical modelling and Southern Ocean dynamics (see *Nature* 401, 515; 1999). In fact, redundancies will occur in ocean/shelf dynamics. The centre is continuing to support core research in numerical modelling and Southern Ocean dynamics.

Tough measures bring a scarred science back to the world stage

Ten years after the fall of the Berlin Wall, east German scientists are putting the pain of reunification behind them, as Alison Abbott reports.

It took Gottfried Geiler 21 years to be awarded the title of professor after gaining the required qualifications in 1961. At the time, most of his contemporaries at the University of Leipzig — at least, those prepared to join the Communist party — made the academic grade within five years.

Geiler's career as a research pathologist was held back because, as a Christian, he never accepted Communist ideology. He was, however, a competent scientist who nurtured 52 PhD students and belonged to the prestigious scientific academy, the Leopoldina.

But by the time the Berlin Wall fell on 9 November 1989 — an event referred to in Germany as the *Wende* — shortages in the German Democratic Republic (DDR), including research materials, had become critical, the state's control on education and research had tightened to a stranglehold, and it was extremely difficult to maintain the quality of research.

"When the wall fell, the working lives of east German scientists were transformed: it was the difference between night and day," says Geiler. Most important, he says, was the fact that east German scientists could for the first time travel freely to meet their colleagues in the west.

New, better institutes

Although such freedoms had a social price, Geiler, like many others, believes it has turned out to be justified. Scores of new research institutes, where facilities are often better than those in west Germany, attest to the successful construction of a modern research infrastructure. But the pain of transition has been severe, particularly during the first few years, when decisions were being made about who could stay in research, and who had to go.

In the DDR's time, science was organized along Soviet lines, with research conducted almost exclusively in 60 or so overstuffed institutes of the Academy of Sciences. Little research was carried out in universities; the free-thinking required of researchers was not considered by the Communist party to be compatible with teaching the young.

In preparation for the country's reunification, the then West and East German governments signed a joint communiqué in July 1990 calling for "a single research landscape", based on the structures of west Germany. In



Hands across the great divide: the reorganization of east Germany's research institutes has not been without errors and resentments, but the benefits are now emerging.

practice, this meant bringing research in the east back into universities, which were to be governed by the west German university law, and bringing former academy research institutes under the umbrella of existing west German research bodies.

These are the Helmholtz Society, which runs national research centres; the Max Planck Society (MPS) and the Fraunhofer Society, which run basic and applied research institutes respectively; and the Leibniz Society, a collection of institutes which do not fit other categories.

Some saw this arrangement as an act of colonization, not only bringing the benefits of the politically independent and well-funded western organizations, but also imposing their faults. These are widely acknowledged to include, for example, administrative inflexibility and permanent contracts for most senior academic staff.

The two governments asked the Wissenschaftsrat, Germany's advisory science council, to evaluate the east German academy's research institutes and recommend how many of their 24,000 posts — around 40

per cent of which were for scientists — should be recreated, and in what sort of institute.

The scheme that failed

The Wissenschaftsrat recommended that just over 11,000 posts (just over half for scientists) should be kept in new research institutes, distributed more or less evenly around the 'new *Länder*' (states). The founding committees of these new institutes were to hire staff through open competition.

In general the founding committees used the standard selection criteria used in the west, such as publication record. They also tended to ignore a broad recommendation from the Wissenschaftsrat that about 10 per cent of posts at each level should be given to west German scientists. This recommendation had been made partly to bring experience of western-style competitive research to east German groups, but also to ensure that the best scientists from the east would be able to stay in the system; in the event, many more west Germans were taken on at higher levels.

The Wissenschaftsrat said that posts should be created for a further 2,000 hand-picked scientists to help re-establish research within east German universities. These became the so-called WIP scientists, named after the fund, the *Wissenschaftler-Integrationsprogramm*, set up to keep them working until universities were in a financial position to hire them permanently.

Lives of east German scientists were transformed

But this scheme eventually failed, mainly because universities, also grossly overstaffed, had to shed their academic staff by the thousand, and were reluctant to earmark new posts for WIP scientists. Only about 350 WIP scientists have found permanent jobs in universities, according to a survey by GEW, the union for education and research staff. But as the statistics do not distinguish between permanent, temporary, full- and part-time staff, even this is probably an overestimate.

The failure of the WIP programme has been the most disappointing part of the restructuring of east German science, says Wilhelm Krull, the prime organizer of the Wissenschaftsrat's research evaluations in the early 1990s, and now head of the Volkswagen Foundation: "It meant that the hoped-for boost for research in the universities did not happen as planned."

The universities themselves took one or two years longer than the research institutes

to restructure themselves. After reunification, all staff were checked by local committees for their professional competence and 'personal integrity' — which usually meant whether they had worked for the secret police (Stasi), or used Communist party contacts to the detriment of other staff.

Losers in the competition

Even those who passed through this process could not be sure of holding their jobs. Thousands of positions were eliminated as university departments were closed or slimmed down to avoid duplication with similar departments nearby, and many scientists failed to win the open competitions to retain their jobs.

Many east Germans still complain that teaching skills, their major strength, were not taken adequately into account. West Germans were quick to make the most of the new academic opportunities in the east at a time

Table 1 **Origins of senior staff at Leibniz Society institutes**

	East	West	Foreign	Total
Institute Directors	10 (23%)	32 (74%)	1 (2%)	43
Department Heads	95 (58%)	59 (36%)	9 (6%)	163
Total	105 (51%)	91 (44%)	10 (5%)	206

*Some institutes have more than one institute director. Some institute directors are also department heads, but are only recorded as institute directors.

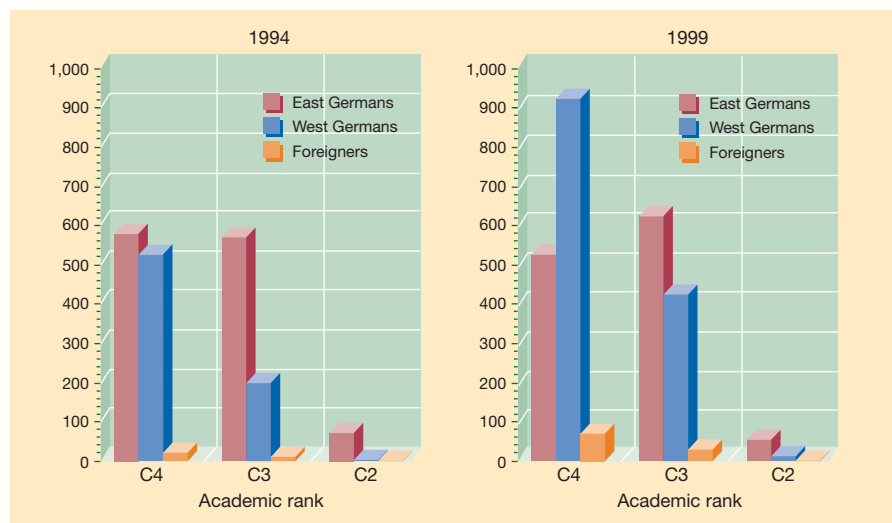
when few such positions were available at home. "One of the hardest times of my life was when I had to sign 1,500 dismissal notices in one day," says Peter Gutjahr-Löser, chancellor of the University of Leipzig.

Little is known about the fate of the tens of thousands of university staff, including nearly 1,000 professors and 14,000 *Mittelbau* (lower ranking academics), who lost their jobs. "No one talks any more about the fate of these people, most of whom were very highly qualified," says Dagmar Schipanski, recently appointed minister of research in Thüringen, east Germany, who was president of the Wissenschaftsrat between 1996 and 1998, and stood in the federal presidential elections last year.

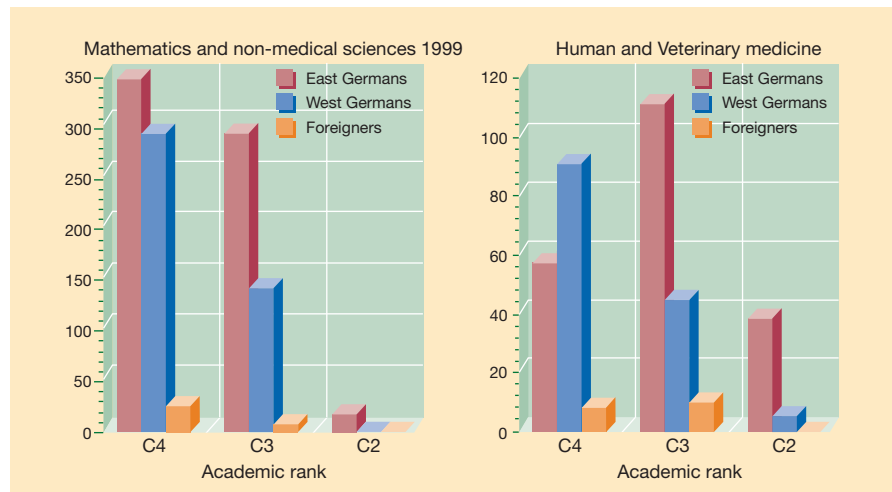
Only the state of Saxony has assembled nearly complete records of what happened to its own 18,000 non-clinical university staff who lost their jobs, out of the total of 28,000 employed in 1989. Around 1,000 were dismissed because they were considered politically unsuitable, and an additional 830 resigned spontaneously. Twenty-three per cent settled for a 'payoff', 16 per cent went into retirement or accepted early retirement, and just over 40 per cent had to go because their positions were cancelled.

Many university staff felt they had been unfairly dismissed and took their cases to court. In Saxony, 1,859 dismissed academics took legal action; 130 won their cases and 82 were reappointed. But the process took its toll. At least one suicide has been recorded as a result of the dismissals, that of a Leipzig professor of cell biology who successfully appealed against his dismissal for using his Communist party links to the detriment of his colleagues. He was not reinstated because his post had been filled during the protracted court hearings (see *Nature* 378, 530; 1995).

A survey by *Nature* (see the adjacent graphs) shows that only one-third of the highest ranking professorships (C4) in east Germany are now occupied by east Germans, a significant decrease since 1994, when not all the new positions had been filled. Lower academic positions (C2), nearly all of which were occupied by east Germans in 1994, have also lost ground to westerners, who now hold around 20 per cent of these posts. East Germans have done better than this average in the non-clinical sciences, retaining half of the C4 professorships and two-thirds of the C3 professorships.



Overall staffing figures in east Germany show that senior (C4) professorships have become dominated by west Germans since 1994. West Germans are also making inroads in lower academic positions.



Staffing figures for the non-clinical sciences in the east show east German academics holding a higher proportion of senior positions than west Germans and foreigners.

Within the east German research institutes of the Leibniz Society, three-quarters of institute directors, and over a third of department heads, come from west Germany. The directors of the three new national research centres are west Germans, and 55 per cent of department heads are from west Germany with a further eight per cent coming from abroad.

Even more extreme ratios exist in the 20 Max Planck institutes, with only three of the 240 institute directors and department heads being east Germans. In contrast to universities and other research organizations, 40 per cent of these top jobs are occupied by foreigners. Three of the society's 19 independent research groups are east German; the rest are west German. However, the MPS did not take over existing institutes, but founded new ones from scratch (see below).

Most see the west German 'takeover' as having been inevitable. Benno Parthier, one of only two Academy of Sciences research directors who were never members of the Communist party, finds the survey results "unsurprising". "You have to accept that there will be a lost generation," which disap-

Most see the West German 'takeover' as having been inevitable

peared for the good of east Germany's scientific future, he says. At his own Institute for Plant Biochemistry, from which he retired as director at the end of 1997, "even those hired on temporary grant money come increasingly from west Germany".

Parthier, currently president of the Leopoldina Academy, says the restructuring of the academy "involved few genuine victims of injustice". But Jens Reich, a bioinformaticist originally from the academy's prestigious Central Institute for Molecular Biology (CIMB) in east Berlin, believes the restructuring of the non-university research institutes could have been much fairer. In 1994 Reich was appointed head of bioinformatics at the Max Delbrück Centre for Molecular Medicine, which took over the CIMB.

"We were not treated unfairly, according to western rules," he says. "But the rules were against us. For example, the selection process was in English, whereas we could have done better in Russian, and publication record was a major criterion, whereas we had had few chances to publish in western journals."

There were also cultural differences. "We all spoke German, yet after 40 years of cultural divide it was hard to really talk to each other," says Horst Franz Kern, dean of science at the University of Marburg, who chaired the Wissenschaftsrat's committee on biology and medicine at the time of the *Wende*.

Joint Wissenschaftsrat evaluation committees of east and west German scientists became essentially run by west Germans — and according to west German rules — says another west German member of the committee, because "the east Germans were not used to our open way of discussion and at the beginning were mostly silent". More importantly, the founding committees of the new institutes that made the final recruitment decisions were heavily dominated by west Germans, says Kern.

Max Planck Society's careful planning reaps benefits

Munich

Shortly after the *Wende*, as Germans call the fall of the Berlin Wall in November 1989, the Max Planck Society (MPS) made itself unpopular by refusing to join other research organizations in their rush to rescue good scientists by taking over their research institutes.

The society wanted to stick to its principle — the so-called Harnack principle — that institutes should be created in scientific areas that balance its overall portfolio, and that each institute should be built around the research of prominent scientists recruited from around the world. At the time the society ran about 60 institutes in west Germany.

The society mitigated some of the criticisms it received by creating, at a total cost of DM200 million, 27 temporary research groups within universities, each headed by an east German scientist. The universities hosting the groups agreed to take over the support of each professor and at least part of his or her research group after the end of the MPS's five-year funding period.

Opening new territory

Drawing up clear contracts with the universities made all the difference to the success of this programme, whose aim was to help reintegrate research into universities. All the professors were taken



Markl: new institutes and fresh research.

into faculty, in contrast to the government-sponsored WIP programme, which had the same aim, but was less successful (see page 635). In parallel, the society planned more thoughtfully the directions of 20 new institutes. "We chose themes that extended our research areas — for example, for the first time we set up an engineering institute — and took the opportunity to create institutes in particularly exciting new areas of research," says Hubert Markl, MPS president.

"But in a few cases we continued some of the very good research of the DDR, for example the microstructure-physics work in Halle, in innovative directions."

The MPS ventured into areas of research — such as cognition and evolutionary anthropology — that would have been unthinkable for Germans in the decades after the Second World War, with sensitivities to the abuse of biology very high. Demographic research was also a particularly sensitive area into which the MPS ventured.

"It is perhaps helpful that the directors of our new Institute for Demographic

Research in Rostock are foreigners, and better able to lead discussions which we can't avoid anyway," says Markl.

The proportion of foreign research directors in MPS institutes in east Germany — 40 per cent of the total — is twice as high as in the west. This was a deliberate move, says Markl, "to signal that we do not represent the colonizing power of Wessies [west Germans], but an opening up of the scientific future in Germany as a whole".

The MPS also took pains to consider new approaches to research, putting together groups from different scientific areas in the same institute to promote interdisciplinary working. The Max Planck Institute for Evolutionary Anthropology in Leipzig has departments for developmental and



Pääbo: neighbours spark new interests.

comparative psychology, evolutionary genetics, linguistics and primatology, for example.

"It's hard to say how far physical proximity will help generate genuinely new interdisciplinary ideas," says Svante Pääbo, one of the institute's research directors. "But already my interest in chimp evolutionary genetics has been sparked." ■

Although many argue that the speed of the evaluations led to injustices for individuals, Kern defends it as having been necessary to minimize the uncertainty for east German researchers. In fact, speed turned out to be politically expedient. Money from the federal government to create and furnish new institutes flowed freely at first, but by the mid-1990s the party was over.

"If the restructuring had been extended by even a year or two, the new research institutes would have found themselves a lot poorer," says Gerhard Neuweiler, a professor of zoology at the University of Munich who was a member of the Wissenschaftsrat's evaluation committee. Initially the research ministry earmarked DM2.4 billion (US\$1.3 billion) to build up research infrastructure in the new *Länder*, but such subsidies were sharply reduced in the mid-1990s (see *Nature* 376, 543; 1995).

Age strata a future problem

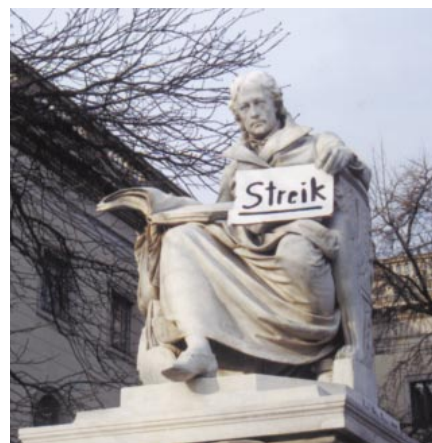
Some individual recruitment mistakes may have been made at the beginning, says Kern, and the "high visibility of the rapid restructuring procedure made it look like a *Blitzkrieg*". Some scientists, including Dieter Simon, president of the Wissenschaftsrat at the time of the *Wende*, have claimed that in the early days of recruitment some opportunistic west Germans were appointed as university professors by naive recruitment

committees, even though they were not from the top ranks.

It is impossible to assess this claim, but concern that second-class appointments were made has certainly bred resentment. Pavel Ströhmer, a former academy researcher in the DDR who joined a biotechnology company after the *Wende*, says that good east Germans were "frequently" not recognized and passed over "for west Germans who were more stupid".

But Krull says that, "given the overall size of the venture", the reorganization seems to have been a general success scientifically. A new round of evaluations of most of the research institutes, again conducted by the Wissenschaftsrat, is now drawing to a close. Its final conclusions will not be reported until next year. But informal reports from scientists taking part in site visits have been positive, says Krull.

The challenge for east Germany now is to maintain the growth of scientific success despite the relative poverty of the region. Already a serious problem is looming for the universities. In the early 1990s, many of the tenured professorships were awarded to scientists in their late forties and early fifties, and older staff were encouraged to take early retirement. But the result is what Gutjahr-Löser describes as a "middle-aged stronghold," with few new positions likely to open up for younger academics for many years.



Humboldt called for teaching alongside research, but east German universities are still unsettled.

A parallel problem was created by the rapid expansion of universities in west Germany in the 1970s. Many of the academics given tenure in that period will retire within the next five years, opening up what many east German universities fear will be irresistible opportunities for the best of the younger west German professors, who have proved their research skills in the east, to return to the west.

The funding gap still counts

Although this will in turn open up some new positions in the east, it will also directly

Rush rescue programme has achieved mixed results

Munich

When the Wissenschaftsrat, Germany's science council, was called in to redesign east Germany's scientific landscape after reunification, its prime aim was to rescue the best of the DDR's science. At the same time, it also wanted to make the most of a unique opportunity to explore novel avenues of science.

According to Wilhelm Krull, who organized the Wissenschaftsrat's evaluations in 1990 and 1991, there was a great eagerness to allow new research directions — particularly those with an interdisciplinary perspective — to blossom in the east and to fill in some of Germany's scientific gaps. "To some extent this was achieved," says Krull.

But the Wissenschaftsrat had only one year to visit all the 60 or so institutes of the DDR Academy of Sciences and make recommendations about each one's fate. As a result of this rush to get things moving, hopes of initiating new research areas not covered in the west, or to abandon research considered 'unnecessary', have frequently evaporated.

"Given that our mandate was to rescue

the best existing science," says Gerhard Neuweiler, a zoologist from the University of Munich who took part in the evaluations, a concerted effort to engineer the content of east Germany's research landscape "became impractical".

Risks and uncertainties

Nor was there always consensus about what might constitute 'unnecessary' research, says Neuweiler. For example, the Wissenschaftsrat eventually recommended that research on zoo animals in Berlin should be taken over by a new Institute for Zoo and Wild Animal Research, despite the feelings of some scientists that such research should be a low priority for Germany.

Some research institutes, such as the Institute for Neurobiology (IFN) in Magdeburg, did successfully move in new directions. Research at the IFN is widely considered to be of a high international standard. The risk taken by Henning Scheich, its director, in moving from a safe chair in west Germany to establish a systems-based, highly interdisciplinary institute focusing on learning and memory, seems to have paid off. The number of

high-impact publications emerging from the institute has rocketed.

Another clear success has been the Astrophysical Institute (AIP) in Potsdam. Its director, Günther Hasinger, has added an experimental dimension to the strong theoretical research programme previously carried out at the DDR institute on which the AIP was founded.

Hasinger has also given the institute some long-term security by creating a role in developing instruments for national and international astronomy projects, thus helping to anchor the institute in the international astrophysics community.

Other institutes have had more difficulty getting off the ground. The Institute of Molecular Pharmacology (FMP) in east Berlin had great trouble attracting a director. The Wissenschaftsrat recommended that it focus on addiction research, which is not strong in Germany. However, this idea was not taken up by the founding committee, which was dominated by cell biologists who were unenthusiastic about building on the behavioural pharmacology strengths of the former academy institute.

increase the difficulty of filling them with good candidates. Research institutes and universities already experience great difficulties with recruitment, and worry about being able to maintain their new research strengths, let alone build on them.

One reason is that family life in east Germany remains less comfortable than in the west. Another is that, outside Berlin, salaries of all state employees, including academics and researchers, remain nominally 86 per cent of those in west Germany. A far higher differential was agreed by the government at the time of reunification to reflect differences in prosperity and living costs between east and west. The differential was expected to close within a few years, but this has been delayed by the disappointing development of the east German economy.

Many institute directors and university rectors have argued for finance ministries in the new *Länder* to make extra funding available for academics, on the grounds that competition for academic posts — unlike that for most public service positions — is at a national, or even international, level. But their requests have fallen on deaf ears.

Moreover, the combination of idealism and opportunism that led many west Germans eastwards is no longer a driving force. "Along with the lower salaries, this makes it almost impossible for us to fill vacancies," says Bernhard Sabel, director of the Institute

of Medical Psychology at the University of Magdeburg.

The problem is compounded by a significant decline since 1992 in the numbers of scientists in east Germany completing *Habilitation* — the qualification required in German-speaking countries to become a professor. According to Josef Lange, general secretary of the German University Rectors' Conference, this reflects the loss of confidence of east Germans in an academic career following the years of uncertainty in universities and research institutes in the early 1990s. These institutions are, therefore, likely to remain somewhat dependent on recruitment from the west for several years.

State finance ministries make exceptions to the '86 percent rule' for some top professorships. Even so, says Gutjahr-Löser, it is very hard in east Germany to come up with good enough packages to attract top scientists from elsewhere, particularly as they would be unlikely to persuade their research teams to take a cut in salary to accompany them.

Gutjahr-Löser is still smarting from the last-minute loss of a prominent west German molecular biologist who had been made "a very generous offer" by the research ministry in Saxony to set up a well-equipped research laboratory, but then took a position at a west German research institute which was able to offer an additional DM750,000.

Institutes experience great difficulties with recruitment already

The poor performance of the east German economy also threatens the investment needed to maintain research laboratories at the internationally competitive level they have in many cases reached thanks to federal government subsidies.

But, whatever the problems of the past few years, it is clear that research is now much healthier in east Germany than in other former Communist countries in central and eastern Europe. "This difference should not be played down, despite the harsh first years which are now behind us," says Reich.

Perhaps most importantly, the personal pain experienced during the turbulent years of the restructuring appears not to have cast a permanent shadow. The new generation of east German researchers seem genuinely unaware of an east-west divide. They work and socialize easily with each other, and consider that they have equal academic chances. ■

This Briefing has been written by Alison Abbott with additional reporting by Eva von Scharper

The first three scientists to be offered the post of FMP director each rejected the terms on which the offer was made, judging the risk of making a success of the remotely located institute to be too high.

Walter Rosenthal, who became the FMP's first director in 1996, has now got the institute on its feet. He has introduced a new research plan, which the Wissenschaftsrat approved last year. This involves identifying new concepts of drug therapy by combining synthetic chemistry (also a strength of the former academy institute), structural biology and molecular pharmacology.

"I am shamelessly reductionist," says Rosenthal, who has also engineered the FMP's participation in the German Human Genome Project. The Institute will move onto the campus of the Max Delbrück Centre for Molecular Medicine, a new national research centre, next year.

Stability and cooperation

Another research centre which got off to a rocky start was the Institute for Molecular Biotechnology in Jena, which had trouble keeping some of its earlier research directors (see *Nature* 385, 761; 1997). But that, too, is now stabilizing.

A major component of the original mission of the institute, defined by its

founding committee, was in the area known as 'evolutionary biotechnology', a new but risky idea for identifying new drug targets through complex changes in gene expression induced by evolutionary pressures.

The institute is now, like the FMP, moving towards a more reductionist approach of identifying single specified genes as potential drug targets, although a



The Einstein Observatory, Potsdam: now used for solar studies by the Astrophysical Institute.

department of evolutionary biotechnology remains.

Whereas most of the non-university research institutes appear to be creating valuable niches for themselves, the situation in the universities seems patchier mainly because they had a much later start. While the DDR Academy of Sciences was dissolved and a fresh start made on 1 January 1992, the universities had to struggle hard, and individually, with their own restructuring.

With less money to play with, the universities have always experienced difficulties in attracting and keeping top scientists. But they are often good at collaborating with neighbouring institutes and concentrating resources in particular areas. The University of Magdeburg, for example, has made neuroscience one of its priorities and has developed strong links with the IFN.

Cooperation between the new institutes and universities is an important feature of east Germany. A high proportion of institute directors in east Germany have a joint position at the university. The institutes usually also play an important role in training research students — unlike the situation in west Germany, where rivalries often hinder collaboration. ■

Conventional crops are the test of GM prejudice

Sir—Millstone *et al.*, in their Commentary in last week's issue¹, claim that 'substantial equivalence', a rule governing toxicity testing of genetically modified (GM) crops, is a pseudo-scientific concept.

One of their arguments is that it is insufficient to test glyphosate-tolerant soybeans (GTSBs) for toxicity and health problems; beans must be tested specifically after glyphosate treatment when isoflavone levels are modified. Millstone *et al.* suppose that toxicity could result from unspecified interactions with the single gene incorporated in the GTSB. The GTSB (Roundup) technology actually requires glyphosate spraying only early in the season when soybean plants are small and weeds a strong competitor for soil and light resources. The beans themselves form months later when the effects of the biodegradable glyphosate sprays have disappeared.

As in all discussion about GM plants, it is important to ascertain the applicability of these arguments to conventionally bred crops, either to avoid, or to expose, simple prejudice against the technology itself. We are unable to think of any environmental stress condition in the quality or supply of light, in the supply of water, minerals or a host of pests and diseases which does not modify isoflavone levels and indeed the content of a host of potential carcinogens that are found in most plants².

Using the logic of Millstone *et al.*, every new crop seed variety would have to be separately tested for toxicity when it has been treated with every herbicide, every pesticide, fertilizer variations, attack by every individual predator, infection with every individual disease and grown in an astronomically large number of different environmental combinations. We would be drowning in toxicity tests. And all these tests would be simply to eliminate the remote possibility that a particular balance of carcinogenic chemicals inside the plant induced by a unique set of conditions might interact in some unexpected way with the many new genes that are combined by conventional plant breeding in the new seed variety.

If this phenomenon ever happens it is more likely to occur in conventional new-variety crops, because many new genes are present rather than the single well characterized *trans* gene and its protein product in a GM plant. Only two examples, to our knowledge, of the environmental induction of a toxic compound that was not detected during routine testing have ever emerged out of the many millions of conventional crop lines produced. Psoralen was found to

accumulate in one line of insect-resistant non-GM celery in response to light, and to cause skin burns³. Cool weather-induced toxic accumulations of solanine caused the withdrawal of the non-GM Magnum Bonum potato line in Sweden⁴.

The UK Health and Safety Executive concluded, after 25 years of intensive scrutiny, that GM food technology is one of the safest yet developed⁵. GM soya has been eaten for 3–4 years by hundreds of millions of people in the United States and Europe with no untoward effects. The type of ill-informed logic expressed by Millstone *et al.* obstructs the acceptance of a new and far safer technology, simply because the authors don't like it. Their arguments are a distraction from the task of developing a sustainable and environmentally friendly agriculture, which combines the best of conventional plant breeding approaches with the new technologies.

Anthony Trewavas*, C. J. Leaver†

**Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK*

†*Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK*

1. Millstone, E., Brunner, E. & Mayer, S. *Nature* **401**, 525–526 (1999).
2. Ames, B. N., Profet, M., & Gold, L. S. *Proc. Natl Acad. Sci. USA* **87**, 7772–7776 (1990).
3. Ames, B. N. & Gold, L. S. in *Fearing Food* (eds Morris, J. & Bate, R.) 18–38 (Butterworth Heinemann, Oxford, 1999).
4. Van Gelder, W. M. J., Vinke, J. H. & Scheffer, J. J. C. *Euphytica* **88**, 147–158 (1988).
5. Wilson, M. A., Hillman, J. R. & Robinson, D. J. in *Fearing Food* (eds Morris, J. & Bate, R.) 58–78 (Butterworth Heinemann, Oxford, 1999).

Substantial equivalence is a useful tool

Sir—We would like to respond to last week's Commentary¹ in which Millstone *et al.* incorrectly assert that: "Substantial equivalence is a pseudo-scientific concept because it is a commercial and political judgement masquerading as if it were scientific. It is, moreover, inherently anti-scientific because it was created primarily to provide an excuse for not requiring biochemical or toxicological tests."

The concept of substantial equivalence was developed proactively before any new genetically modified (GM) foods came to the market. It was first described in an OECD (Organization for Economic Co-operation and Development) publication in 1993^{2,3} produced by about 60 experts from 19 OECD countries, who spent more than two years discussing how to assess the safety of GM foods. Most of these experts, all nominated by governments, were regulatory scientists from government agencies and ministries responsible for consumer safety.

In 1996, participants at an expert World Health Organization/Food and Agriculture Organization consultation⁴ recommended

that "safety assessment based upon the concept of substantial equivalence be applied in establishing the safety of foods and food components derived from genetically modified organisms". This represented an endorsement by experts based on three years' experience in the safety assessment of various GM foods.

Substantial equivalence is not a substitute for a safety assessment. It is a guiding principle which is a useful tool for regulatory scientists engaged in safety assessments. It stresses that an assessment should show that a GM variety is as safe as its traditional counterparts. In this approach, differences may be identified for further scrutiny, which can involve nutritional, toxicological and immunological testing. The approach allows regulators to focus on the differences in a new variety and therefore on safety concerns of critical importance. Biochemical and toxicological tests are certainly not precluded.

Since the concept of substantial equivalence was first described, several new foods have been assessed and knowledge has accumulated on how to use the concept. In parallel, the OECD, its governments and others have continued to review its adequacy in food safety assessment and to develop supporting tools⁵. The OECD's task force on the safety of novel foods and feeds (chaired by P. M.), in particular, continues to focus on the application of the concept. This includes work on assessment methodologies when substantial equivalence cannot be applied, as well as efforts to identify the critical nutrients and toxicants found in major crop plants, as a focus for the demonstration of substantial equivalence.

More than a decade of work by the OECD and its member governments was recognized by the heads of state and government of the G8 countries when they met in June in Cologne and invited the OECD task force to undertake a study of the implications of biotechnology and other aspects of food safety. This additional challenge is certain to lead to further reflections on the concept of substantial equivalence.

Peter Kearns*, Paul Mayers†

**Organization for Economic Co-operation and Development, 2 rue Andre-Pascal, 75775 Paris Cedex 16, France*

†*Health Canada, Postal Locator 2203G3, Tunney's Pasture, Ottawa, Ontario K1A 0L2, Canada*

1. Millstone, E. *et al.* *Nature* **401**, 525–526 (1999).
2. *Safety Evaluation of Foods Derived Through Modern Biotechnology: Concepts and Principles* (OECD, Paris, 1993).
3. *Food Safety Evaluation* (OECD, Paris, 1996).
4. *Biotechnology and Food Safety* (FAO/WHO, Rome, 1996).
5. <http://www.oecd.org/ehs/food/index.htm>

No GM conspiracy

Sir—Last week's Commentary by Millstone *et al.*¹ is misleading and inaccurate. The authors do not seem to be

aware of a meeting organized by the Organization for Economic Co-operation and Development, held in Aussois, France, in March 1997, which is about to be published. Hence they present an outdated view of the use of substantial equivalence, while their characterization of this tool as the outcome of an international conspiracy to foist genetically modified (GM) foods on a gullible public is beyond belief. Do Millstone *et al.* really believe in a worldwide conspiracy? They have no evidence for this assertion.

Second, by such accusations, Millstone *et al.* denigrate the whole regulatory process and all the hard-working academics who make up the British regulatory committees. Throughout their article, they reveal their ignorance of the way the regulatory process works. Substantial equivalence is a tool only: the 'first cut' at the decision-making rather than a quick solution, as the authors infer. I was chairman of the Advisory Committee on Novel Foods and Processes (ACNFP) from 1989 to 1997, and we never received any political or commercial pressure when making decisions. I totally reject the slur on our integrity.

Finally, Millstone *et al.* are wrong in what they say about GM soya. They imply that a herbicide that has no effect on the target enzyme, and that does not persist in the plant, has far-reaching effects on intermediary metabolism. If the herbicide does not affect its target, how can it affect the plant? Not only is this idea bizarre, it is wrong: GM beans have been analysed after treatment with herbicide and their composition is unaffected²⁻⁴. Millstone *et al.* are also wrong to say that treatment with herbicide alters the isoflavone content; they do quote one paper reporting some variation but ignore others showing that this difference is well within the normal range of isoflavone content⁴. More than 1,400 compositional analyses of Roundup Ready soya beans have been conducted, showing that there are no significant differences in the key soya bean nutrients and anti-nutrients; these data have been reviewed by the ACNFP. The authors use the literature selectively to make their point, which is not good enough for any scientific journal.

I contend that these data establish that GM soya beans are as safe as conventional soya beans. This conclusion was reached by the ACNFP after a thorough safety assessment, using substantial equivalence as a key safety assessment approach. This conclusion has been confirmed by regulatory agencies in the 13 countries that have approved these GM soya beans. This food has been used commercially for four years, and 300 million Americans are currently eating it with no sign of a problem.

How did such a mish-mash of old hat sociology and poor science get published? I would like an assurance from the editor that all such contributions — especially from activists — are rigorously refereed. *Nature*, in my view, damages its reputation by publishing such propaganda.

Derek Burke

13 Pretoria Road, Cambridge CB4 1HD, UK

1. Millstone, E., Brunner, E. & Mayer, S. *Nature* **401**, 525–526 (1999).
2. ACNFP Annual Report Appendix XII, par. 8 (1998).
3. Padgett *et al.* *J. Nutr.* **126**, 702–716 (1996).
4. Taylor *et al.* *J. Agric. Food Chem.* (in the press).

● The publication policy of Commentaries is long established. The Commentary section, like the Book Reviews and Correspondence sections, is intended to convey original and stimulating opinions, both solicited and unsolicited, from outside contributors. Publication does not imply endorsement by *Nature* of the authors' views. In contrast to Articles, Letters and Brief Communications, not all Commentaries are refereed, but it is *Nature's* policy to seek advice especially where we believe there is a risk of readers being misled. On the few occasions when, despite that policy, falsehoods or significant misrepresentations have been published, rebuttals or corrections have been published as soon as possible. — Editor, *Nature*

Putting transparency into ethical balance

Sir— Science nowadays often responds to a particular ethical challenge in isolation from others, so that any overall philosophical perspective might not be apparent to the public. This can contribute to an impression that scientists muddle through on the defensive when it would be fairer and more persuasive if they displayed the relationships and consistencies between ethical positions on different matters. How might it be possible to achieve this?

People make ethical decisions by weighing opposing arguments, pressures and concerns, drawing on their feelings about the issues as well as on their thinking. Is earning a profit from drug or seed sales less ethical than providing these commodities cheaply or free to the needy? Is whistleblowing in public better than a quiet warning? Should the integrity of experimental animals take priority over prospects for medical advances?

Most people would probably agree that the answer to each of these questions is sometimes yes and sometimes no. Judgements are personal and specific to the cir-

cumstances. When people disagree it is usually because they assign different weights to particular factors, rather than differing on what the considerations should be.

I believe that most people could agree on generic principles which cover most or all issues in current contention, deep though their differences might be over interpretation and application. I suggest that these three would suffice:

(1) Fair shares of burdens and fruits. Goods and services for one population should not bring undue disadvantage to another, for example in the form of chemical, radioactive or noise pollution. Areas of risk include genetically modified crops; information technology and its use (on whom and for whom); databases and copyright law and intellectual property law for private versus public benefit, among others. Some inequities arise through omission, especially of the application of science to agriculture, health and education in the poorer countries (a theme of UNESCO's recent World Conference on Science).

(2) Obedience to truth. Honesty over details must be reconciled with perspectives on what is important in academic and corporate science and their applications. This should apply to the use, abuse and fabrication of data; commitment to establishing and acknowledging the whole truth; and the ethical tensions that arise from private research funding in public institutions.

(3) Respect for life. The uniqueness and integrity of all life forms and the environments that support them must be protected, whether as elements of the planet's gene pool or as individuals with conscious lives. This will expose conflicts between the needs of one organism and another, but although conflicts have existed throughout biological evolution, it is inevitable that we now accept the responsibility of resolving them by human decision.

Each of these principles is formulated as a balance to reveal why people think and feel differently about given problems, and even why they change between related issues. Ethical positions will and should always shift at the levels both of individuals and of society as a whole, and consensus will not and should never be complete or stable. We need challenge to recognize danger and test responsibility, and we also need to understand the patterns in different judgements to develop the democratic mandate for science to go forward to meet the needs and aspirations of society. Engagement with critics of science needs to shift from tactics to strategy; namely from whether in the short term new technologies are to be developed, to how in the long term they are to be certified and introduced.

Dai Rees

European Science Foundation, 1 quai Lezay-Marnésia, 67080 Strasbourg Cedex, France

Do-it-yourself climate prediction

Anyone with a home PC could join climate modellers in their attempt to forecast how the Earth's climate will evolve in the next century.

Myles Allen

The tortuous process of drafting, reviewing and revising the third assessment report of the Intergovernmental Panel on Climate Change (IPCC) is well under way, following the meeting of chapter authors in Arusha, Tanzania, on 1–3 September. Even those of us only peripherally involved have been burning the midnight oil to complete analyses 'in time for IPCC', and this is nothing compared with the heroic efforts of the lead authors.

Only the most hardened cynic would deny the value of this five-yearly stimulus to the climate research community, and the usefulness of the reports it generates. One of the IPCC's greatest achievements has been to bring national governments 'on board'. Nevertheless, any such intergovernmental process sits uneasily in the era of stakeholders and focus groups. Large numbers of scientists gather together periodically and attempt to forge a consensus about the nature and scale of the problem of global warming. This is followed by a gathering of an even larger number of policy-makers, primarily politicians and civil servants, to decide what to do about it. A bemused public must then be persuaded, for its own good, to go along with the solution.

Reaching the public

As long as politicians only propose targets that they think they can sell to a generally indifferent electorate, it seems unlikely that we will achieve the order-of-magnitude reductions in greenhouse-gas emissions required to make a significant impact on the problem. And as long as climate research remains confined to large, centralized institutions, this indifference is likely to remain.

Most of our understanding of climate change is still based on physical principles and the simulations of climate models. The emergence of a detectable signal is a welcome new piece of data for model validation, but it does not fundamentally change our understanding of the problem. Moreover, the signal means little to the public, being based on changes in temperature that seem insignificant and which occur on scales too large to matter. Without an unambiguous indicator of climate change with which to capture the public imagination, as the Antarctic ozone hole did in the 1980s, the gap between specialist and public opinion will only widen.

Effective communication of uncertainty is notoriously difficult, and the analysis of uncertainty in climate forecasts remains rudimentary. Modellers do, of course,

examine how sensitive their predictions are to changes in model parameters, but only one or two parameters can be varied at a time. It is the subtle interactions between errors that we must analyse. Another approach is to compare results from different climate models, but the few models available share assumptions, algorithms and, sometimes, even source code. They provide, at best, an incomplete estimate of uncertainty.

Perturbing parameters

In an ensemble weather-forecasting system, the 'best-guess' initial state is deliberately perturbed in various directions to establish the range of tomorrow's weather that is consistent, at a given level of confidence, with today's observations. In an ensemble climate forecast, perturbations would also have to be applied to all uncertain parameters in the climate model. Unlike weather forecasters, who have several methods of selecting 'optimal' perturbations, climate researchers have no option but to perturb parameters, within their ranges of uncertainty, and integrate the model to see what happens.

Almost all such perturbed models would soon drift into an unrealistic climate. Thus, for a 50-year forecast, we would first need to perform large numbers of simulations of the past 50 years, both with and without external influences such as increasing greenhouse gases. We could then discard all those perturbed models that were either inconsistent with the observed record or inconsistent with our (much less certain) estimate of what the twentieth century would have been like in the absence of external influences. Provided the perturbations are large enough for us to be able to discard most models, there are objective grounds for believing that the survivors would span the current range of uncertainty. These could then be integrated onwards to provide a probabilistic climate forecast.

Hundreds of thousands of 50-year integrations would be required to explore all the relevant parameter combinations in a full-scale climate model. This is well beyond the capacity of current and planned modelling facilities. Or is it? The 'HadCM3' model from the UK Meteorological Office is one of the latest generation of coupled models used in the IPCC2000 Third Assessment Report. A single 50-year integration would take about six months on a reasonably up-to-date home personal computer, while the memory and disk-storage requirements are no larger than those of an advanced computer game.

More than a million volunteers are

currently scanning radio-telescope data on their home PCs for signs of extraterrestrial intelligence (see *Nature* **400**, 804; 1999). Could a similar number be recruited for the more practical, albeit more demanding, task of forecasting the climate of the twenty-first century? After all, it is possible that none of the participants in the SETI@home project will find signals of alien life. But, provided we keep the records, someone could definitely tell their grandchildren that it was they who, on a US\$1,650 PC, made the most accurate forecast of the global mean temperature of 2050.

Although the range of predictions would be large, the chances are that only a small fraction of these perturbed models would reproduce the recent warming and show no significant change over the coming 50 years (unless, of course, there is something crucial that we haven't thought of yet). This in itself would be helpful for counteracting the powerful lobby still promoting the notion that there is nothing to worry about. Perturbing parameters is only the beginning; the next step would be to feed in new parametrizations, or even, as PCs improve, higher-resolution models.

We would still, of course, need to make small numbers of integrations of very high-resolution models at supercomputing centres. But the ability to perform million-member ensemble simulations, even at a relatively coarse resolution, would profoundly affect climate modelling. Models would no longer be fixed representations of the climate, revised with great effort only every few years. Instead, they would be probabilistic, fuzzy entities, reflecting our fuzzy understanding of the climate system, in which tunable parameters are represented by uncertainty ranges rather than by single 'best-guess' values.

Get the children involved

Anyone respectable enough to sit on a peer-review committee will probably find the idea of getting schoolchildren to run full-scale climate models on their parents' PCs completely daft. But there will be others for whom the idea is as natural as Amazon.com. If one of those happens to be chief executive officer of a major PC manufacturer or software house with an urge to save the planet, perhaps they could get in touch.

Myles Allen is in the Space Science Department, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK.

e-mail: m.r.allen@rl.ac.uk

If you have access to a PC and would like to participate in a fuzzy-climate modelling initiative, please register on <http://www.climate-dynamics.rl.ac.uk>

Save Mozart for later

Neurobiology offers no evidence that infancy is a privileged time for learning.

The Myth of the First Three Years: A New Understanding of Early Brain Development and Lifelong Learning

by John T. Bruer

Free Press: 1999. 244 pp. \$25

Elizabeth Spelke

The media have widely reported that research on brain development singles out the first three years of life as a time of great plasticity, when young children have opportunities for learning and development that are closed to them later on. On the heels of this message come other messages for parents and policy-makers: that infancy and early childhood are the best time to expose one's child to Mozart, French, full-time parenting or a designer preschool, and that it is the most effective target age for interventions to improve disadvantaged children's cognitive potential and minimize their risk of emotional and social maladjustments. John Bruer looks at the evidence behind these messages and concludes that the emperor has no clothes. Nothing that has been learned about human brain development justifies a special focus on the first three years of life or supports any general prescriptions about the treatment of young children.

Writing for a

balanced review of three strains of research on brain development: studies of synapse formation and loss in the cerebral cortex of humans and other primates; studies of critical periods in the development of patterns of cortical connectivity in cats and monkeys; and studies of the effects of environmental complexity on the brains of rats. In no area, he argues, do research findings show that the potential to learn declines after the first years of human life.

Although the number of synapses in the cortex grows enormously over human infancy, no evidence relates this growth to variations in learning capacity, which appears to increase with the decline in synapse numbers that occurs gradually throughout childhood. Although some cognitive systems, particularly those subserving aspects of visual function and language, show critical periods in development, when experience has large effects on function, not all critical periods happen in the first three years, nor do all cognitive systems show such effects, and, for those that do, the experiences that alter the course of brain development typically lie far outside the normal range. And although laboratory rats that live in more interesting environments have more synapses in some parts of their brains than do those that live in barren ones, the effects of environmental

neuroscientists, cognitive scientists and policy-makers. Why were the implications of research on brain development in infancy so badly misunderstood by so many spokespeople? Do studies of children's developing brains have any implications for studies of their developing minds? Can such studies inform child-rearing and educational practices in any useful ways?

Bruer considers these questions in his book, and treated them in depth in a thoughtful essay that preceded it ("Education and the brain: A bridge too far", *Educational Researcher* 26, 4–16; 1997). He concludes that a crucial gap must be bridged before research in neuroscience can inform efforts to foster children's learning: the child's developing cognitive systems must be understood at a functional level, and cognitive development must be related systematically both to developing brain function and to educational practice.

Research in developmental cognitive science has revealed a good deal about what children know and how they learn, and it has constructed an array of tools for extending that knowledge. Those who turn to findings in brain science for insights into children's development ignore this evidence and these tools at their peril. Suppose, contrary to fact, that research in neuroscience had revealed an early critical period for the development of a part of the brain involved in music processing in the adult. What should parents do with this information? Should they sing more to their babies? Play them Mozart minuets? Begin instrumental music lessons? Any of these suggestions might hold intuitive appeal, but neuroscience and intuition do not combine to yield sound educational practice. Understanding the implications of research on brain development requires systematic study of what children know and how they learn in different cognitive domains and at different times.

Bruer makes these points but does not emphasize them for an obvious reason: if zero to three is not a critical period for brain development, then we need not worry about what we should do if it were. It would be unfortunate, however, if readers concluded from the debunking of the myth of zero to three that all attempts to link brain development, cognition and child-rearing are misguided. Bruer's own professional life suggests otherwise. As president of the James S. McDonnell Foundation in St Louis, he has been centrally involved in fostering research connecting neuroscience to cognition and cognitive science to education. *The Myth of the First Three Years* shows us



Early learning by academic design: The *Teletubbies* television series was developed for preschool children in consultation with child psychologists.

how not to make those connections, but the most important book that Bruer could give us will tell a different story. This book ends where his work, and the emerging field of developmental cognitive neuroscience, begin. ■

Elizabeth Spelke is in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Oiling the wheels of controversy

The Deep Hot Biosphere

by Thomas Gold

Springer: 1999. 235 pp. \$27, £19

R. John Parkes

Modern society relies heavily on petroleum for energy, and life would be very different without it. The term 'fossil fuel' effectively informs us that the origin of oil is from the bodies of dead plants and animals, altered initially by microbial decay, and then 'cooked' under pressure and at relatively high temperatures during burial to several kilometres' depth.

We have 'harvested' this deep resource extremely rapidly, but we also realize that we will eventually run out of this precious commodity. But what if someone told you that this was all wrong and that the hydrocarbons that make up petroleum are constantly refilling reservoirs. Interested? Well, you should read this book. Thomas Gold postulates that oil and gas are of primordial, abiological origin, part of the solid material that accreted to form the planet some 4.5 billion years ago. As the planet subsequently heated up (through radioactive decay and gravitational compression and sorting), only partial melting occurred, producing the different layers of the Earth (crust, mantle and core), and enabling the primordial hydrocarbons and/or their precursors to survive. Gold suggests that these are concentrated at great depths — 100 to 300 kilometres — and are constantly moving towards the surface as fluids and gases via conduits within rocks and collecting in oil reservoirs.

As you might expect, this is a very controversial theory, with both passionate supporters and objectors. It is also in conflict with the basis for current petroleum exploration. It was first presented in the late 1970s, when it was considered that oil reserves would last for only another 15 years, prompting an oil crisis. Not surprisingly, as the theory promised continued supplies of hydrocarbons, it received considerable attention from petroleum entrepreneurs, but it was, and still is, broadly rejected by scientists. An additional aspect of the theory

is that oil and gas should not be restricted to sedimentary deposits, as dictated by the biogenic theory of oil formation, but should be widespread in all rock types, including rocks formed from magma.

This would greatly widen the areas for oil exploration, and in 1986 Gold persuaded the Swedish government to drill 6.7 kilometres into granite to prove his theory. The results of this experiment are reported in detail, but unfortunately, little of this has been published in the scientific literature, which is essential to validate a theory of such major significance. Gold explains, with some frustration, that this is due to scientists' unwillingness to be open to this new theory, leading them to reject his results. If commercial quantities of oil and gas had been produced, Gold's theory would have been willingly accepted; however, only limited amounts of oil were produced. Gold claims that this is sufficient to prove his theory, but scepticism prevails.

One major problem is that petroleum contains 'molecular fossils', clear fingerprints of a biological origin, and thus cannot be of primordial, abiological origin. Gold, however, addresses this problem by hypothesizing that, as the hydrocarbons migrate upwards, they are intercepted by a deep subsurface bacterial biosphere, which uses some of the hydrocarbons as an energy source. Hence, it is these bacteria that provide the biological fingerprint to oil and not organisms that originally lived at the surface and gained their energy from sunlight via photosynthesis.

Gold suggests that this deep biosphere may extend down to 10 kilometres, which would mean high temperatures (between 150 and 300 °C), and hence he has termed it 'the deep, hot biosphere'. The current temperature maximum for bacterial life is 113 °C; hence, bacteria living at even higher temperatures would have to exist to occupy this zone, and 300 °C seems improbably high. Pressures at this depth would also become limiting. In addition, hydrocarbons are relatively resistant to bacterial degradation, especially under the anaerobic conditions that must prevail in the deep subsurface. Therefore, it would be very difficult for bacteria to thrive under the conditions outlined by Gold.

Although many of the molecular fossils in petroleum could be explained by an origin from subsurface bacteria, some are clear indicators of surface photosynthetic organisms (for example, porphyrins and dinosteranes) and hence of a biological origin for the oil. Another problem is that petroleum is made up of isotopically light carbon, which again is indicative of a biological origin. Not necessarily so, says Gold, because he suggests that, during migration to the surface, isotopic fractionation of hydrocarbons containing isotopically heavy carbon would also

result in isotopically light hydrocarbons being more abundant in oil reservoirs. He also states that some commercial methane sources are much lighter isotopically than can be explained by cell debris from plant photosynthesis, but this ignores the fact that methane-forming bacteria produce an extra isotope fractionation. Gold's deep-Earth gas theory, which he extends in this book to include black coal and diamond formation, the concentration of metals and involvement in earthquakes, is so radical that there are many inconsistencies and contradictions with empirical observations. The justification for these extensions may in places prove far from satisfying and sometimes also too anecdotal. Nevertheless, the book makes interesting reading, as Gold presents his evidence skilfully. You may not agree with him, but you have to appreciate his fresh and comprehensive approach to these major areas of Earth science.

At least one area Gold discusses in his book is starting to be substantiated, but there is no evidence that this is being fed by very deep primordial hydrocarbons. Surprisingly large bacterial populations have been found at considerable depths (up to a few kilometres) in a range of different subsurface environments, including terrestrial aquifers, granites and basalts, Cretaceous shales, marine sediments and the rocks beneath. There are even some locations where bacterial activity and populations increase in deeper zones and where geospheric processes may be involved in fuelling a bacterial biosphere to much greater depths than we had previously thought possible. Unfortunately, these are not fully reviewed in the book. The magnitude of this biosphere may also be huge, as Gold predicted. A startling example of this is the discovery that some North Sea oil reservoirs produce 3–16 kilograms of high-temperature bacteria along with the production fluid each day. In his book, Gold considers the implications of subsurface life for the origin of life on Earth and its subsequent evolution, and finally the possibility of life on other planets.

The book is clearly one man's view of major processes in the Earth sciences, and demonstrates that scientific debate is alive and well. Science is not just about facts and data, but is hypothesis-led and thrives on controversy — and few people are more controversial than Thomas Gold. ■

R. John Parkes is in the Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK.

More specialized extremes

Enigmatic Microorganisms and Life in Extreme Environments

edited by Joseph Seckbach

Kluwer Academic, £210, \$360

Much ado about some thing

The Nothing That Is: A Natural History of Zero

by Robert Kaplan

Allen Lane: 1999. 246 pp. £12.99

Ivor Grattan-Guinness

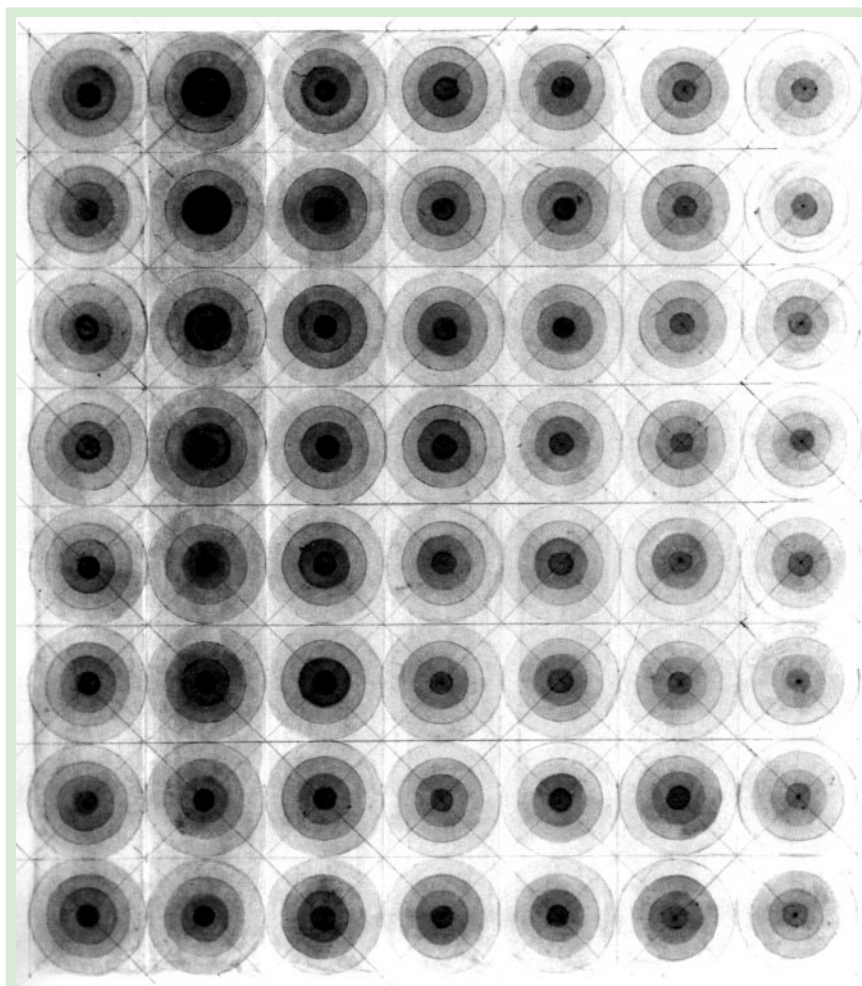
This popular book deals with a fascinating aspect of mathematics: the occurrence of zero within the arithmetics of integers and real numbers, and its many representations as numerals. Robert Kaplan is at his best on the latter: a wide range of examples is given, both of numerals and also of possible origins for the signs '0' and 'o'. On this last point he gives more credit to the Greeks than do other authors, but as the history is so obscure no definitive position can be maintained. Kaplan is careful to stress that Indian mathematicians were the first to recognize that zero could act as a pukka number, and not only as a space filler to distinguish, say, 34 from 304 and 3,004, a practice that is at least Babylonian.

The author also covers related questions, such as possible values, if any, for $0/0$. He discusses the role of zero elsewhere in mathematics, for example with infinitesimals. Probability theory is omitted, however, although there are interesting questions about non-contradictory statements that nevertheless have value zero in a probability calculus.

On number systems, the main point of contrast is with roman numerals, where Kaplan rather exaggerates the difficulties of calculating in that system. He could also have made this contrast for us: that if today is Friday, then two days ago it was Wednesday, because we count today as the zeroth day; the the Roman equivalent was in calling Wednesday three days earlier.

The author notes the concern with 'nothing' in some of William Shakespeare's plays. The general implementation of the Hindu-Arabic numeral system occurred in Britain around Shakespeare's lifetime, and the playwright would not have missed the resemblance of '0' to the 'Wooden O' of his theatre building. Among modern examples of the cultural history of zero not noted here is John Cage's notorious musical composition *Four Minutes, Thirty-Three Seconds*, in which any number of musicians play nothing for this length of time. In lasting 273 seconds, the title surely alludes to -273° Celsius, the absolute zero of temperature. Another apparent candidate is Samuel Beckett's play *Waiting for Godot*, that is, God-O = God-zero = no God, one of the play's main themes.

The book is likely to be widely read; but its success is somewhat limited by several



Zeroing in on process

Working in the late 1960s and early 1970s, a group of American artists rejected formal procedures in art. The result — process art — was possibly the most ephemeral movement in art history; which seems appropriate for art that was centred more on the act than on the final product. The untitled work above, by Eva Hesse, is featured in

the exhibition "Afterimage: drawing through process" (catalogue published in association with MIT Press, \$35, £24.50). The exhibition, which was organized by Cornelia Butler, has recently ended at the Museum of Contemporary Art in Los Angeles. It will be at the Contemporary Arts Museum, Houston, Texas, next year.

considerations. Surely, the author missed a rare opportunity to explain the mistiming of the obsession with the end of the millennium. It does not happen this December, because when the Christian calendar system was instituted in the sixth century, no year zero was imposed; it is, I believe, unique as an application of the positive and negative integers in which zero is missing. The reason seems to have been misunderstanding zero to be nothing, a cultural distrust that was then related to views such as nature abhorring a vacuum. This is a fine example with which to emphasize that zero is some thing of a kind (for example, $7 + 0 = 7$, whereas $7 + \text{nothing}$ is not defined), but the author makes little of it. In general, the difference between zero and nothing is

not sufficiently emphasized.

Another of the book's limitations is in its treatment of negative numbers, which could have been given more attention. Like zero, negative numbers have been treated with mystification and distrust over the centuries, even though they have natural interpretations, such as directed distances (+5 feet forwards, -5 feet backwards) and economic debt and credit (arguably, the origin of the sign '-' in the Late Middle Ages). Issues over the legitimacy of zero often were associated.

Kaplan frequently deploys nice analogies instead of explaining relevant features. For example, when describing the creation of the calculi by Isaac Newton and G. W. Leibniz, he correctly stresses the differences but does not explain the source of them,

which lies in changing dimension. If one takes an infinite sequence of decreasing lines, then in the limit, one arrives at a point, and so changes from dimension 1 to dimension 0. Newton followed this practice in defining his 'fluxion', although unclearly. By contrast, in Leibniz's theory, one adds to a given variable line x an infinitesimally short 'differential' incremental line dx , with dx as its own increment, and so on; limits are explicitly avoided, and ' dy/dx ' is the ratio, usually finite, of two differentials. This role of dimension could have been explained; instead, we learn about "the free play of mind, which browses on the pastures of phenomena", and find quotations from a novel and from Meister Eckhart.

More should have been said about set theory. For example, in the summary of non-standard analysis as the modern theory of infinitesimals, there is a poem by Christian Morgenstern instead of a relevant explanation. Set theory has an empty set $\emptyset$, which is also not 'nothing': the author notes the definition of zero as the unit set of $\emptyset$, but he does not discuss the related tri-distinction between $\emptyset$, zero and nothing, which Bertrand Russell effected around 1900 (after anticipation by Gottlob Frege).

A similar non-nothing worth noting, and omitted in Kaplan's book, is the empty space, such as that between words. Their absence or presence is essential in contexts such as 'READINGANCIENTPAPYRI' and using modern computer programmes. Also, the bibliography is available only on the World-Wide Web. This is a regrettable feature, for the book as it stands is useless, since no details can be checked or pursued in it.

Thus, *The Nothing That Is* is somewhat frustrating. The book as a whole is less than the sum of its parts — to invoke another fascinating mathematical topic with a long history, which has a zero and even a 0/0 of its own with algebraic logicians such as George Boole and C. S. Peirce. ■

Ivor Grattan-Guinness in the Middlesex University Business School, Enfield, Middlesex EN3 4SF, UK.

More mathematical somethings Discovering the Truth and Beauty of the Cosmos: A Mathematical Mystery Tour

by A. K. Dewdney
Wiley, £17.99, \$22.95

The Mathematics of Measurement: A Critical History

by John J. Roche
Athlone, £55

Mathematical Sorcery: Revealing the Secrets of Numbers

by Calvin C. Clawson
Perseus, \$26.95

Science in culture



Design by digits

The Collier Campbell painted designs

Martin Kemp

The origins of the term 'digital', which has come to signify the ultimate in depersonalized processing, lie in that most specifically human of instruments, the hand. The original 'digits' of computation were those ten fingers on which rudimentary acts of calculation were first performed. Now, in the age of computers, practitioners in almost every area of visual representation in art and science have witnessed severe polarization between designs arising from hand-eye coordination and the visual products of technology. The life's work of the remarkable design partnership, Collier Campbell (the archive of which is soon to be sold), proves that the divorce between the handmade and the machine-produced is not irrevocable.

Entering the world of textiles in the 1960s from engagement in painting rather than training in design techniques, the sisters Susan Collier and Sarah Campbell had a mission. It was to replace the tediously repetitive, graphically lifeless and coloristically crude textile patterns then generally available on the mass market with prints that retained the inherent life of freehand paintings. The motive was at once aesthetic and social, since they aspired to bring the costly and elite qualities of the individually made item within the reach of the widest possible public.

The painted designs that poured forth in a torrent of creative enthusiasm, in teeming various styles, possess a vigour of mark-making, vitality of line and energy of colour more akin to a Matisse painting than to a stock 'repeat'. Indeed, they set out to "cheat the repeat", that is to say, to arrive at designs which, if repeated above and below, and on either side of one another, would flow dynamically across the whole surface, without the monotony of insistent repetitiveness.

Beginning commercial production with Liberty of London Prints in 1961, Collier (joined in 1968 by her sister) pushed printed textiles into new technical and visual realms. They insisted that printers exploit and extend their technologies to translate the painterly freshness

and colorism of their designs into prints on a variety of fabrics, so that they could

be reproduced without degradation of effect. The premise on which they worked was that the eye can unerringly identify something that expressed the reach of hand, and that our aesthetic response to a hand-made item has its own unique character, something at once visual and somatic.

They were relying, without theory, on how the eye can undertake awesomely subtle acts of minute perceptual discrimination. These are the acts that allow us to recognize thousands of individual faces even from a fleeting glimpse of a mobile expression; to identify instantly a friend's handwriting on an envelope; and to distinguish a Leonardo da Vinci drawing from a pupil's pious copy. The tiny shadings of difference involved in such human discriminations are beyond description, relying upon what Collier Campbell term "inking without words".

After almost 40 years of undiminished artistic fertility, working with leading international companies and fashion designers, Collier Campbell find themselves in a changed world. The present danger is, as they say, that "the hand can't be seen; only the price". The problem lies with the increasing domination of manufacturing and retail by a few conglomerates. The tendency has been to concentrate on the 'real estate' value of a few 'designer names' which are used to endorse a huge range of products, from perfumes to pants.

The core of the matter is not therefore a necessary incompatibility between hand and technology, between digits and digital, but how we can establish modes of practice that permit the visual traces of human presence to speak eloquently through the new technologies. The final product does need to be literally handmade, but if all we can offer in the popular market-place are mechanical designs devoid of human vibrancy, we will be impoverishing our visual culture. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Revolution in the ocean

Victor Hensen realized that in the sea the very small feed the very large.

Victor Smetacek

Plants dominate life on land, but in the sea they are outweighed by animals. The ocean contains less than one per cent of plant biomass, but the proportion for animals is much larger. Yet the annual production of organic matter is about the same on land and in water. The realization that the ocean's animals are fed by a thin soup of minute algae — the phytoplankton — did not emerge gradually but was the inspiration of one man, Victor Hensen, who published his revolutionary thoughts in 1887.

The study of planktonic organisms had begun 50 years earlier, but scientists believed that the food supply for all the ocean's animals was provided by rivers and coastal vegetation. They seemed unable to conceive that the small could feed the large. Hensen's thinking was revolutionary in both senses of the word: he thought of planktonic populations as rapidly revolving links in a food-chain leading from the very small to the very large and he applied this conceptual framework, which he called the "metabolism of the sea", to the quantitative study of ocean biology. This upset received wisdom, and launched a new discipline. The ridicule heaped upon him by leading scientists of the day testifies to the novelty of his thinking.

Hensen, a professor of medical physiology at Kiel University, was drawn to the study of the sea by the plight of the German fisheries, reeling under fluctuating fish stocks.

His aim was to provide fisheries management with a true scientific basis; to him the clear first step was a reliable estimate of the "yield of the sea". Since direct assessment of fish stocks was not feasible, Hensen reasoned that the distribution and abundance of fish

eggs, provided that they were dispersed uniformly by the currents, should provide a measure of population size. Hensen observed dispersion rates by following glass floats, and convinced himself that the eggs would indeed be distributed uniformly. So he designed some quantitative sampling methods and started counting eggs.

A net designed to sample fish eggs will also collect plankton, and Hensen's vision must have emerged in the many hours spent searching for eggs in thickets of plankton under the microscope.

Unlike his predecessors, he looked beyond the trees, saw the forest, and developed his grand view of how oceans function. The time was ripe for a quantitative investigation of the primary food (*Urnahrung*) of the marine animals.

Hensen braved the choppy Baltic in small boats and counted plankton in rough weather. His first publication on the sub-

ject was a 105-page monograph intended to present methods and preliminary results. In it he coins the term plankton — derived from Homer's *Odyssey* and implying aimless drifting. Hensen begins with the statement that plankton, as well as being "of interest for its systematics and anatomy", was "without doubt of great importance for the entire metabolism of the sea". In the last few pages he compared his estimates of plankton production per square metre with that of agricultural fields and conjectured that plankton was equally, if not more, productive.

We now know that Hensen's conjecture was too high by a factor of five. Further, his methods should have given much lower estimates, because his nets collected only a fraction of the total plankton. Added to this, phytoplankton blooms (when his catches were large) were restricted to only a number of weeks in the year. Clearly he had made up his mind before acquiring the data. The school established by Hensen at Kiel

University proved the importance of plankton and based modern biological and chemical oceanography on an agricultural model.

Although Hensen's belief in the uniform distribution of plankton and the adequacy of his methods was disproved by his colleagues, he defended them right up to his last paper, published posthumously in 1926. He was forced onto the defensive by biologists working on systematics and anatomy, with Ernst Haeckel, the formidable Darwinian and meticulous draughtsman of plankton (as seen in the drawings in this article), at the forefront.

Haeckel seems to have believed that small is beautiful, but not important, and his mentor Johannes Müller, who launched the study of plankton in the 1830s, jokingly called it "philosophical dirt".

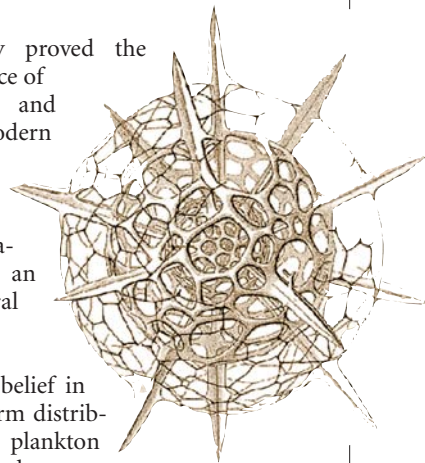
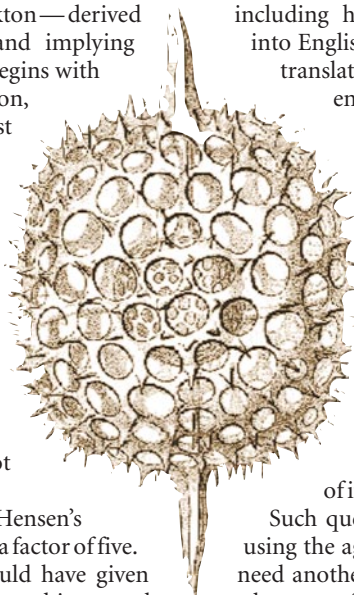
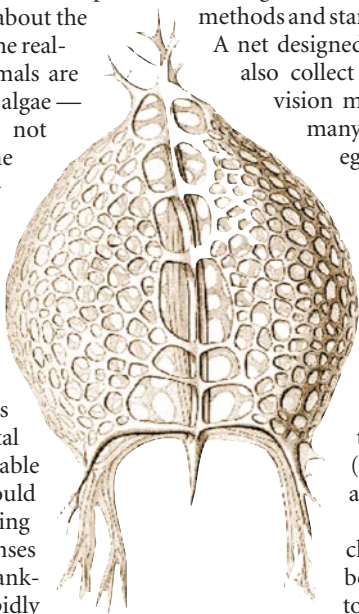
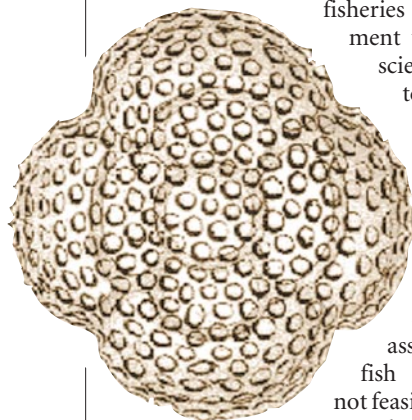
Haeckel attacked Hensen's methods, data and "quantifications". Haeckel's works, including his attacks, were translated into English. Hensen's works were not translated, and Haeckel's name has endured while Hensen's has faded.

Fisheries still reel from fluctuating fish stocks and the causes are still being debated, although we have a fairly accurate estimate of production from the ocean and of yield from the sea. The ratio of these is about 600:1 — so where is all the production going, and what determines how much of it enters the fish pool?

Such questions cannot be answered using the agricultural model. So do we need another revolution? Whatever the substance of the research in the future, Hensen's revolution will persist: plankton is indeed the provider.

Victor Smetacek is at the Alfred-Wegener-Institute for Marine and Polar Research, 27570 Bremerhaven, Germany.

e-mail: vsmetacek@awi-bremerhaven.de



Hurricane heat engines

H. E. Willoughby

A hurricane's path is quite easy to forecast, but its intensity is not. A new model, which takes into account the interaction of the storm and ocean-surface temperature, gives remarkably accurate retrospective simulations.

On the night of 3–4 October 1995, Hurricane Opal turned into a forecaster's nightmare. This atmospheric vortex, which had wandered around the southern Gulf of Mexico since late September, had gradually strengthened to become a weak hurricane on 2 October. Then, under the influence of a low-pressure trough blowing eastwards over North America, Opal lunged abruptly towards the southern shoreline of the United States. During its northward career, Opal crossed a deep pool of warm water that had detached from the Gulf Stream Loop Current and intensified dramatically. In just 22 hours, its strongest winds increased from 40 to 67 metres per second, taking it from category 2 to category 4 on the Saffir–Simpson hurricane scale (Table 1).

Few coastal residents who stayed up to watch late weather broadcasts worried about the hurricane as they went to bed that night. But the situation had changed by sunrise and, for a few hours, it looked as though there could be a repeat of the 1969 Hurricane Camille. (Camille, the only category-5 hurricane on record to strike the US mainland, killed 256 people.) The complication in 1995 was that Opal came upon them overnight, with little warning and essentially no time to evacuate. But then, even more suddenly than they had intensified, Opal's winds weakened to 50 metres per second before the eye passed onshore at Pensacola Beach, Florida. In the end, Opal took nine lives in the United States and caused an estimated \$3,000 million in property damage. Like last month's Hurricane Floyd, Opal's main effect was inland flooding, which killed 50 people in Mexico and Guatemala during the storm's formative stage.

'Hurricane' is the term used in the Western Hemisphere for the general class of tropical cyclones, including western Pacific typhoons and similar systems, that are known simply as cyclones in the southern

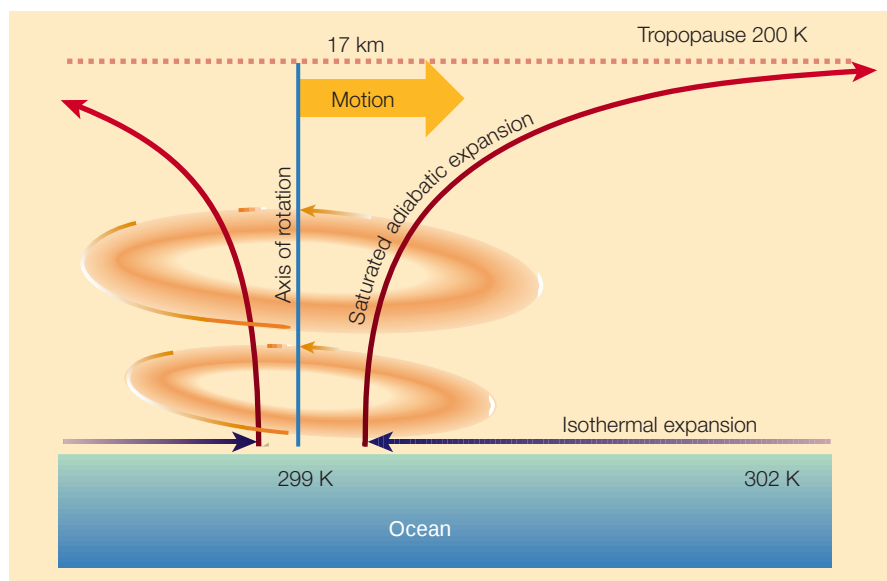


Figure 1 Tropical storms as thermodynamic engines. Air takes up energy, primarily latent heat stored in water vapour, as it spirals into the lower levels of the vortex under the influence of friction. It converges towards the eyewall — a ring of convective clouds that encloses the clear central eye. As the air ascends to the tropopause (the top of the troposphere, where the temperature decreases with height), the vapour condenses, converting the latent heat into sensible heat which is, in turn, converted to mechanical energy that can do work against friction or strengthen the vortex. The energy realized through this cycle is proportional to the difference in temperature between the ocean at roughly 300 K and the upper troposphere at around 200 K. Storm-induced upward mixing of cooler water reduces the ocean-surface temperature by a few degrees, and can have a considerable effect on the fastest wind speed attainable.

Pacific and Indian oceans. They are intense atmospheric vortices that derive their energy from the warm surface waters of the summertime tropical ocean. Their damaging winds are concentrated within about 100 km of their centres, just outside the clear central eye. When a hurricane threatens, everyone wants to know where it will go, so the most widely appreciated success in hurricane forecasting has been in more accurate track predictions. About 80–90% of a hurricane's motion depends on flow in the surrounding, undisturbed atmosphere¹. Predictions of this 'steering flow' have improved in parallel with advances in the art of numerical weather prediction. If present trends continue, by the year 2005 track forecasts made 48 hours in advance will be as good, on average, as those made 24 hours in advance were in 1970.

On the other hand, forecasters claim little skill in predicting the intensity of hurricanes beyond simple extrapolation and climatological experience. For many hurricanes, extrapolation is good enough because they

strengthen or weaken slowly. But in the case of Hurricane Opal, a phenomenon known as 'rapid deepening' occurred — the central, sea-level pressure fell suddenly as winds increased.

Rapid deepening is responsible for most typhoons and all Atlantic hurricanes with winds of greater than 50 metres per second (Saffir–Simpson categories 3, 4 and 5). Although only 20% of hurricanes fall into this class, they cause 80% of hurricane damage in the United States². Moreover, the onset of rapid deepening is unpredictable — it can change a category 2 hurricane into a category 4 or 5 hurricane in less than a day. In other words, the intensities of the most damaging hurricanes are the most difficult to predict. Yet the paper by Kerry Emanuel on page 665 of this issue³ reports startlingly good retrospective simulations of hurricane intensity changes.

The thermodynamics of a hurricane can be modelled as an idealized heat engine⁴, running between a warm heat reservoir —

Table 1 The Saffir–Simpson scale

Category	Pressure (hPa)	Wind (m s ⁻¹)	Damage
1	Above 980	33–43	Minimal
2	979–965	43–50	Moderate
3	964–945	50–59	Extensive
4	944–920	59–69	Extreme
5	Below 920	70 or more	Catastrophic

the sea, at around 300 K — and a cold reservoir, 15–18-km up in the tropical troposphere, at about 200 K (Fig. 1). A hurricane's maximum intensity (that is, the strongest wind speed or lowest sea-level pressure) is proportional to the difference in absolute temperature between these two reservoirs. Nonetheless, most real hurricanes are weaker than predicted by models based on pre-existent sea-surface temperatures.

What Emanuel has done is to apply to hurricanes a simple, time-dependent model written with his student Lars Schade⁵. The model represents the hurricane as a completely circular vortex that moves along a pre-defined track. The model's initial condition is adjusted to match the observed intensity, and the rate at which this intensity changes, at the start of the forecast. Ocean-surface temperature and depth of the warm surface layer are represented on a coarse spatial grid. Turbulent heat exchange at the surface drives the hurricane model. It is proportional to surface wind speed and temperature from the oceanic grid points beneath the storm, averaged on circles around the storm's centre position. Momentum transport from the vortex to the ocean causes the warm surface layer to mix with the cooler water below.

The key to the model's success is the reduced heat supply as the surface water cools in response to the storm. Despite many simplifications, the model produces spectacular forecasts of maximum winds for memorable storms such as Hugo (1989), Andrew

(1992) and Opal. Other hurricanes — such as Camille (1969), Gilbert (1988) and Gloria (1985) — where the forecasts are less successful point to considerable roles for atmospheric interactions or actual ocean-temperature structures that differ from the monthly climatic averages used in Emanuel's preliminary calculations.

In response to the near-disaster caused by Hurricane Opal, a symposium on intensity changes in tropical cyclones was held at the 1997 meeting of the American Meteorological Society in Phoenix, Arizona. There, a consensus emerged that intensity changes in a hurricane are a response to three factors — forcing by the surrounding atmosphere, the oceanic heat source and the storm's internal dynamics. The difficulty of reliable predictions was attributed to the complex interplay of these factors. Emanuel's result, if it stands the test of forecasting experience, will shift the consensus towards a dominance of oceanic forcing. Perhaps intensity forecasting will turn out to be a much less difficult problem than we thought just a few years ago. ■

*H. E. Willoughby is in the Hurricane Research Division of the National Oceanic and Atmospheric Administration's Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, Florida 33149, USA.
e-mail: Hugh.Willoughby@noaa.gov*

1. Holland, G. J. *J. Atmos. Sci.* **40**, 328–342 (1983).
2. Landsea, C. W. & Pielke, R. A. *Weather Forecast.* **13**, 621–631 (1998).
3. Emanuel, K. A. *Nature* **401**, 665–669 (1999).
4. Emanuel, K. A. *Annu. Rev. Fluid Mech.* **23**, 179–196 (1991).
5. Schade, L. R. & Emanuel, K. A. *J. Atmos. Sci.* **56**, 642–651 (1999).

Neurobiology

Straight from the top

Earl K. Miller

Our perceptions and thoughts result from interplay between so-called top-down and bottom-up signals in the brain — bottom-up signals are derived from the physical data that stream into sensory receptors, and top-down signals provide the interpretation of these data. Top-down signals transmit information synthesized from experience, the stored knowledge that provides context and meaning to sensory inputs, so they are central to cognition. But much of what we know about top-down signals is based on indirect evidence, such as their effects on bottom-up sensory processes. In an ingenious experiment presented on page 699 of this issue, Tomita *et al.*¹ provide a direct glimpse of top-down signals in the brain, as well as evidence that these signals emanate, at least in part, from the prefrontal cortex.

To appreciate the roles of top-down and bottom-up signals in perception, take a look at Fig. 1. Bottom-up analyses tell us about the

location, colour and local contours of the spots in the picture. Top-down, experience-dependent processes provide the coherence and meaning, so we perceive that the spots form a familiar figure, a Dalmatian. But top-down signals are not limited to perception. They come into play whenever we use acquired knowledge to direct our attention and determine what things are, how they go together and how we should act. That is, they make complex behaviour possible.

In a normal brain, top-down and bottom-up processes are inextricably linked. Feed-forward projections in the cerebral cortex, which transmit bottom-up information away from sensory receptors, are matched by feedback projections that can transmit top-down signals. This makes it difficult to isolate top-down signals, because most cortical activity in a conscious animal is a combination of top-down and bottom-up processes. Nor has it been easy to find the

source of top-down signals. Each cortical area has widespread connections, and feedback could come from many places. Now, however, Tomita *et al.*¹ have removed the bottom-up signals from the inferior temporal cortex in monkeys, an area of the brain involved in visual memory. What they have left behind are top-down signals from a putative source — the prefrontal cortex.

The prefrontal cortex was a good bet. It is a collection of cortical areas at the anterior end of the brain that is greatly enlarged in primates. Owing to its interconnections with all of the main systems in the forebrain, the prefrontal cortex has long been thought to be involved in synthesizing the data needed for 'executive' control of behaviour². These include 'external' data, such as sensory and motor information, and 'internal' data such as memories and emotional state. Moreover, prefrontal neurons have ideal properties for top-down control — they transmit acquired knowledge.

Work in monkeys has shown that neural activity in the prefrontal cortex can reflect relevant experiences, such as learned associations between diverse stimuli and actions, and even abstract information such as the 'rule' being used to guide behaviour^{3–6}. Although a transient event may trigger this activity, prefrontal neurons can maintain it while the monkey waits for more information or to make a response^{7–9}. Furthermore, there seem to be functional interactions between the prefrontal and inferior temporal cortices, because deactivating the prefrontal cortex decreases neural activity in the inferior temporal cortex¹⁰.

Tomita and colleagues¹ explored the function of prefrontal top-down signals when monkeys recalled visual images stored in the inferior temporal cortex. The appearance of a cue object instructed monkeys to recall, and then choose, another object that was associated with the cue during training. In the intact brain, information is shared between the inferior temporal cortices in the



Figure 1 Example of top-down signals in visual perception. Bottom-up (data-driven) processes analyse the spots whereas top-down (conceptually driven) processes use stored knowledge to provide an interpretation of what they form (a Dalmatian sniffing the ground). If at first you cannot see the dog, you are experiencing bottom-up signals alone.

two cerebral hemispheres. The authors used a surgical technique inspired by Gazzaniga's studies of split-brain patients¹¹ and Mishkin's classic disconnection experiments¹² to sever the connecting fibres (see Fig. 1 on page 699). This meant that each inferior temporal cortex could only 'see' (that is, receive bottom-up inputs from) visual stimuli that appeared in the opposite visual field. The fibres connecting the prefrontal cortices in the two hemispheres were left intact.

When Tomita *et al.* examined the activity of single neurons in an inferior temporal cortex that could not see the cue, this cortex nonetheless reflected the recalled object, albeit with a long latency. The visual information seemed to take a circuitous route, travelling from the opposite inferior temporal cortex (which could see the cue) to the still-connected prefrontal cortices, and then down to the 'blind' inferior temporal cortex. The authors confirmed this pathway by severing the two prefrontal cortices and eliminating the feedback. Sure enough, activity in the inferior temporal cortex was abolished and the monkeys could not complete the task.

Previous studies^{13,14} indicated that the prefrontal cortex is involved in recall, but Tomita and colleagues are the first to demonstrate that interactions between the prefrontal and inferior temporal cortices are directly implicated. Like any great experiment, though, the results raise questions. How ubiquitous is the role of the prefrontal cortex in top-down control? For example, is it also involved when we focus our attention on something we anticipate as being informative? By removing top-down signals, we can examine how they interact with bottom-up signals. Some data indicate that top-down signals directly enhance the activity of sensory neurons¹⁵, but other, less direct, influences are possible. Also, the prefrontal cortex is unlikely to work alone, so the function of other structures will need to be explored.

One of the most exciting things about Tomita and colleagues' study is that it introduces a powerful tool for addressing fundamental questions of cognitive function. Their results also indicate that the prefrontal cortex plays a complicated and central role in brain function. If it exerts such executive control over many brain systems, it is easy to see why damage to the prefrontal cortex produces wide-ranging deficits in functions including memory, attention, selecting behaviours and inhibitory control.

Earl K. Miller is in the Department of Brain and Cognitive Sciences, Center for Learning and Memory, and RIKEN-MIT Neuroscience Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: ekm@ai.mit.edu

1. Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I. & Miyashita, Y. *Nature* **401**, 699–703 (1999).

2. Nauta, W. J. H. *J. Psychiatr. Res.* **8**, 167–187 (1971).

3. Sakagami, M. & Niki, H. *Exp. Brain Res.* **97**, 423–436 (1994).

4. Asaad, W. F., Rainer, G. & Miller, E. K. *Neuron* **21**, 1399–1407 (1998).

5. Hoshi, E., Shima, K. & Tanji, J. *J. Neurophysiol.* **80**, 3392–3397 (1998).

6. White, I. M. & Wise, S. P. *Exp. Brain Res.* **126**, 315–335 (1999).

7. Fuster, J. M., Bauer, R. H. & Jervey, J. P. *Exp. Neurol.* **77**, 679–694 (1982).

8. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. *J. Neurophysiol.* **61**, 331–349 (1989).

9. Miller, E. K., Erickson, C. A. & Desimone, R. *J. Neurosci.* **16**, 5154–5167 (1996).

10. Fuster, J. M., Bauer, R. H. & Jervey, J. P. *Brain Res.* **330**, 299–307 (1985).

11. Gazzaniga, M. *Neuron* **14**, 217–228 (1995).

12. Mishkin, M. *Phil. Trans. R. Soc. Lond. B* **298**, 83–95 (1982).

13. Hasegawa, I., Fukushima, T., Ihara, T. & Miyashita, Y. *Science* **281**, 814–818 (1998).

14. Rainer, G., Rao, S. C. & Miller, E. K. *J. Neurosci.* **19**, 5493–5505 (1999).

15. Desimone, R. & Duncan, J. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).

Quantum physics

Waves, particles and fullerenes

Alastair I. M. Rae

Can the fundamental concepts of quantum physics apply to everyday 'classical' objects as well as those in the atomic and subatomic regime? Can we meaningfully attribute wave properties to an everyday object such as a football, or does quantum theory break down at some level? In one guise or other these questions have been asked ever since the quantum theory was invented.

On page 680 of this issue¹, Arndt *et al.*, a group working under the direction of Anton Zeilinger in Vienna, report experiments showing that molecules of the fullerene C_{60} have wave as well as particle properties, just as predicted by quantum theory. A fullerene

such as C_{60} is of course very much smaller than a football, but it does have a mass at least an order of magnitude greater than that of any other object whose wave properties have been previously observed.

There is very little room for doubt concerning the correctness of the wave-particle duality postulate for fundamental particles. About 80 years ago, Prince Louis de Broglie suggested that 'atomic' particles such as electrons had wave as well as particle properties. This received early confirmation from the experiments of Davisson and Germer², which demonstrated electron diffraction for the first time; and the wave properties of neutrons have been exploited for at

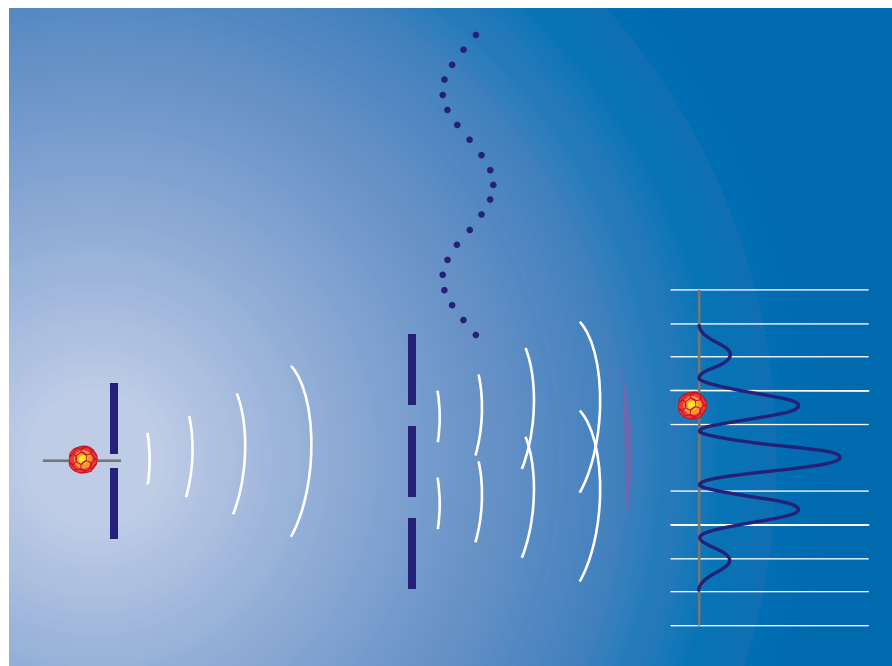


Figure 1 A beam of particles passes through a double slit in the form of a wave and produces an interference pattern. However, whenever the particles are actually detected, their particle properties are manifest. The experiments of Arndt *et al.*¹ demonstrate this effect in the case of C_{60} molecules, which are an order of magnitude heavier than anything previously shown to have wave properties. The molecule may absorb or emit radiation while passing through the apparatus, as indicated by the dotted waveform. Provided the wavelength of the radiation is much larger than the distance between the slits, so that it is impossible in principle to tell which slit the molecule passed through, the interference pattern is unaffected.

least 50 years in neutron-diffraction experiments.

An archetypal example of wave-particle duality is the two-slit experiment (Fig. 1). A wave crosses a screen containing two slits. The wave splits it into two parts, which are re-united on a screen, creating an interference pattern. It is the creation of this interference pattern that is an unambiguous signature of a wave. In particular, the fact that the two components cancel each other out at some points in the pattern is very difficult to explain in any other way. On the other hand, when the object reaches the screen it is always detected as a particle — hence the term ‘wave-particle duality’.

Even on the atomic scale, wave-particle duality can raise difficult questions. For example, does the particle actually go through one or other of the two slits or does it in some way exist in both slits simultaneously? We know that if we put detectors into the slits, we will always find that the particle passes through one or other of them; but an essential consequence of such a measurement is that the interference pattern is no longer formed, so that the wave properties are no longer manifest. Results such as these led Niels Bohr to propose that the type of properties (particle or wave, for example) that we are allowed to attribute to a quantum system depend on the type of observation we make on it. Other solutions to this ‘measurement problem’ have been proposed, including the apparently outrageous ‘many worlds’ interpretation in which such a measurement divides the apparatus and everything that interacts with it into two branches that continue to exist, but are for ever unaware of each other’s existence.

A central question in this debate is if and how quantum theory applies to macroscopic objects. Hence the importance of the paper by Arndt *et al.*¹. For many years Zeilinger’s group has been renowned for its pioneering work in neutron interferometry, but this latest achievement demonstrates the interference of the de Broglie waves associated with molecules of the fullerene C_{60} . A molecular beam emerging from an oven at a temperature of 1,000 K is collimated and passed through a grating whose slits are 50 nm wide and 100 nm apart (the use of a grating containing many slits rather than the two-slit interference discussed above introduces no new issues of principle). The emerging beam is detected and found to form an interference pattern that is completely explicable on the basis that the beam has wave properties.

C_{60} is of course not a macroscopic object, but it is an order of magnitude more massive than anything else that has been studied in this way before. Being a large molecule at a high temperature, its component atoms are in continual motion and it is an essential requirement for successful interference that these motions remain coherent while the

molecule is passing through the slits. Arndt *et al.* point out that this coherence would be destroyed by any process that would in principle allow an observer to determine through which slit the molecule passed. They calculate that C_{60} at this temperature would be expected to emit two or three infrared photons during its passage through the apparatus. But because the wavelength of the radiation associated with these photons is much greater than the distance between neighbouring slits, they carry no information about which slit the molecule passed through and therefore the interference pattern is not affected.

These experiments extend the applicability of wave-particle duality by about one order of magnitude in the macroscopic direction. However there are about 15 orders of magnitude to go before we reach the mass of anything we would normally think of as macroscopic. Some experiments in other fields, such as superconductivity and magnetism (see ref. 3 for review), claim to be demonstrating the quantum properties of

objects with a close-to-macroscopic number of degrees of freedom, but these results are quite controversial. In contrast, the C_{60} experiment is quite unambiguous. If it could ever be extended into the genuinely macroscopic regime, the consequences would be profound. If everything can behave like a wave or a particle depending on how it is observed, what does this mean if it includes what we would normally call observers or measuring apparatus, or even the Universe as a whole? If, on the other hand, this is shown to be impossible in principle, we will have reached the point where quantum physics no longer fully applies and the limitations of the most successful theory in the history of physics will have been found. ■

Alastair I. M. Rae is in the School of Physics and Astronomy, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.

e-mail: a.i.m.rae@bham.ac.uk

1. Arndt, M. *et al.* *Nature* **401**, 680–682 (1999).

2. Davisson, C. J. & Germer, L. H. *Nature* **119**, 558–560 (1927).

3. Chudnovsky, E. M. & Tejada, J. *Cambridge Studies of Magnetism* Vol. 4 (Cambridge Univ. Press, 1998).

Ecology

Power behind diversity’s throne

Shahid Naeem

On average, plants contain less than half a gram of carbon per square centimetre. Yet this thin veneer of living matter, a sort of ‘green slime’ sandwiched between a 100-km-deep lithosphere and a 100-km-high atmosphere, manages to cycle 60 gigatonnes (60×10^{15} g) of carbon per year between the biosphere, lithosphere and atmosphere¹. Clearly, the Earth’s biota has a staggering capability to affect our environment. But this perspective overlooks a critical feature of plant life; this green slime consists of more than a quarter of a million species. What, if any, is the role of such extraordinary diversity? Is it important to the rates of ecosystem processes? Is it important to the stability of ecosystems? We must answer these questions if we are to understand how widespread losses in biodiversity² are likely to affect ecosystem properties and the services they provide³. A rapidly growing body of ecological research deals with this issue, and considerable excitement and controversy surround its findings⁴. However, on page 691 Sankaran and McNaughton⁵ present results from a very different kind of study that challenges both the nature and the direction of this new field.

Although ranging widely in character and design⁶ — from small bottles of microbes⁷ to the BIODDEPTH experiments that span nine European countries — the flurry of recent experimental work on the role of biodiversity in ecosystem processes

has centred on a single line of investigation. Diversity has been manipulated by constructing replicate ecosystems in which some species are excluded from some replicates. This mimics the biodiversity loss that might arise because of stochastic processes, habitat modification, overexploitation, displacement by invasive species or other processes that result in the local extirpation of species⁸.

Although local extirpation of species is perhaps the most visible and disturbing cause of contemporary declines in biodiversity, Sankaran and McNaughton examine whether large-scale variation in diversity is actually caused by factors such as climate, the frequency of disturbance, soil fertility and other factors known as extrinsic determinants of diversity. That is, unlike most experiments, which measure ecosystem properties across a gradient in diversity generated by extirpation, these authors measured ecosystem response to perturbation across an existing gradient of diversity generated by extrinsic determinants. Specifically, they selected three types of savanna grassland in India’s Western Ghat Mountains that varied in diversity due to extrinsic determinants (most likely disturbance), then perturbed replicate plots by clipping, burning or both. The authors measured resistance stability — how much a system changes after perturbation — a widely studied and important property of ecosystems.



100 YEARS AGO

THE BEST EDUCATION FOR AN ENGINEER

A study of the classics and a public school education are frequently regarded as synonymous, and so the advantages of the one are confounded with the advantages of the other. At the present time, when so much attention is devoted in secondary and technical schools to *matter* rather than to *manner*, when the aim apparently is to turn out scientific encyclopaedias rather than fairly well-informed people with cultivated manners, the following opinion expressed by Sir Andrew Noble should be taken to heart by every engineering student: "Speaking as an employer of labour, I should say that we find a pleasant speech and manner, tact in dealing with others, and some power of organisation of the utmost value; and it is precisely those qualities which a boy acquires, or ought to acquire, in his *later* years at a public school. Without such qualities even the highest scientific attainments will never make a captain of industry, and in selecting candidates for appointments the man of business distinctly prefers a youth who has had the benefit of some years at a good school."

From *Nature* 12 October 1899.

50 YEARS AGO

In a fascinating article which appeared in *Nature* of February 6, under the title "Stellar Evolution and the Expanding Universe", Mr. F. Hoyle has brought forward convincing arguments for the permanent creation of matter in space. Many physicists will find it difficult to accept this hypothesis. For if there is any law which has withstood all changes and revolutions in physics, it is the law of conservation of energy, which according to Einstein's formula $E = mc^2$ is equivalent to the conservation of mass. The same strange conclusion has, during recent years, been formulated by Prof. Pascual Jordan, but with an important modification, whereby the conservation law is not violated. This is achieved by taking account of the loss of gravitational energy connected with the creation of particles. As Jordan's papers do not seem to be known to many English-speaking physicists, I have asked him to write a short report of his work, and the following article is a translation of his article made by my collaborator, Dr. H. S. Green.

MAX BORN

From *Nature* 15 October 1949.

One might think that if the end result is the same, with ecosystem properties being measured against a gradient in biodiversity, nothing new should emerge from this line of investigation. Sankaran and McNaughton⁵ found, however, that two measures of resistance stability — stability of species composition and stability of species turnover — showed significant and opposite associations with diversity across, but not within, community types (see their Fig. 1 on page 691). These results show that variation in diversity owing to extrinsic determinants may dominate the relationship between biodiversity and ecosystem properties, whereas variation in diversity within communities may play a substantially smaller or different role than current research might lead us to believe. The authors outline several cautionary notes concerning the importance of distinguishing between the various sources of variation in diversity that emerge from this study.

Although Sankaran and McNaughton's study is experimental, it is not, as the authors acknowledge, free from the difficulties associated with studies of natural gradients in diversity. There remains the nagging concern that neither factor measured (plant diversity or proneness to disturbance) is actually responsible for the patterns observed. Multiple regression analyses support the authors' claims, but these statistical methods merely apportion response variance to measured factors, and can neither identify causation nor rule out the possibility that some covarying but unmeasured factor is the responsible agent. This inability to identify causation or mechanism is the main reason that experimental ecologists avoid studying natural gradients in diversity⁹. For the authors to get around these problems, they would have to

repeat this enormous experiment, manipulating within-community diversity by adding or removing species in replicate plots, and manipulating grass cover to change proneness to disturbance.

Pursuing such a full factorial experiment would quadruple the size of the study, but would probably not change its message. Sankaran and McNaughton⁵ warn us that the effects of extrinsic determinants on large-scale patterns in diversity and the factors, such as local extirpation, that affect fine-scale variation in diversity, are quite different from one another. Understanding the ecological significance of the Earth's extraordinary diversity will require studying both sources of variation. After all, extrinsic determinants, such as atmospheric carbon dioxide, nitrogen deposition and global climate, are themselves changing, much as the local extirpation of species is on the rise. Predicting the ecosystem consequences of biodiversity loss is going to require some branching out from the kinds of experiment that currently dominate research in this area.

Shahid Naeem is in the Department of Zoology, University of Washington, Seattle, Washington 98195, USA.

e-mail: naeems@u.washington.edu

1. Schlesinger, W. H. *Biogeochemistry* 2nd edn (Academic, San Diego, 1997).
2. Pimm, S. L., Russel, G. J., Gittleman, J. L. & Brooks, T. M. *Science* **269**, 347–350 (1995).
3. Daily, G. C. in *Nature's Services* 392 (Island Press, Washington DC, 1997).
4. Grime, J. P. *Science* **277**, 1260–1261 (1997).
5. Sankaran, M. & McNaughton, S. J. *Nature* **401**, 691–693 (1999).
6. Allison, G. W. *Am. Nat.* **153**, 26–45 (1999).
7. McGrady-Steed, J., Harris, P. M. & Morin, P. J. *Nature* **390**, 162–165 (1997).
8. Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. *BioScience* **48**, 607–615 (1998).
9. Tilman, D. *et al. Science* **278**, 1866–1867 (1997).

Astronomy

Super photon counters

John C. Mather

The perfect photon detector would measure the arrival time, the energy, the polarization and the position of every arriving quantum of light, but that is easier said than done. Two groups have now succeeded in measuring everything but the polarization emitted by the Crab nebula pulsar, as reported in *The Astrophysical Journal* by Romani *et al.*¹ and in *Astronomy and Astrophysics* by Perryman *et al.*². Both groups use superconducting detectors to achieve the necessary speed and sensitivity. Making better detectors is far cheaper than building bigger telescopes and for many astronomers having access to the latest electronic detector is the key to progress.

Three years ago, Peacock *et al.* reported^{3,4} that they had detected single optical photons with a superconducting tunnel junction (STJ). A tunnel junction consists of two conductors separated by a tiny gap of insulating material or even a vacuum. If the gap is thin enough, electrons can tunnel across anyway, and if the conductors are superconductors, the junction displays very useful quantum-mechanical properties and electrical nonlinearities. An arriving photon breaks apart the pairs of electrons responsible for the superconducting state, which can then be collected. The main advantages of STJs is that they operate at high speed at very low temperatures, dissipate very little power and are very small.

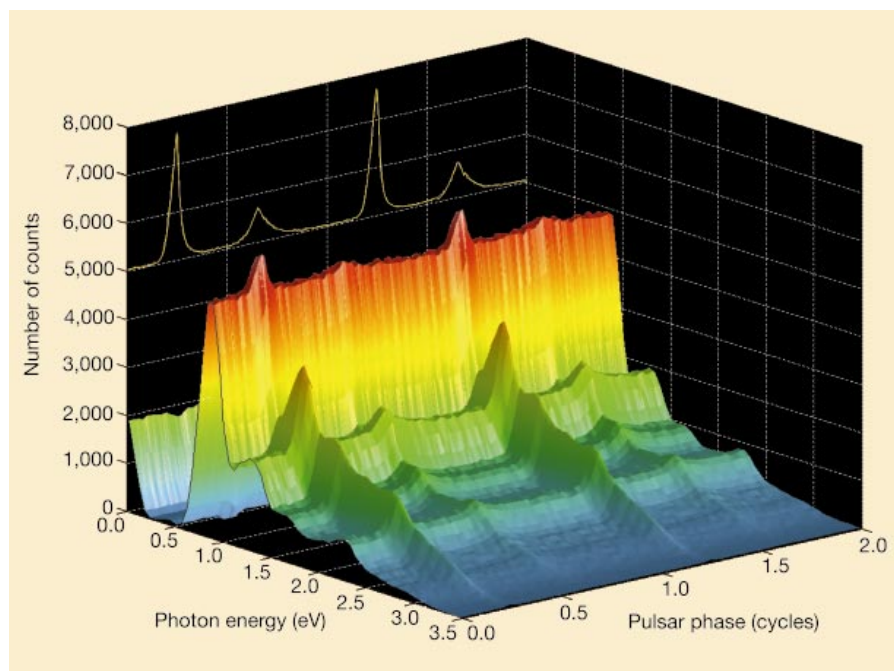


Figure 1 How to catch photons from a spinning pulsar. Time-resolved spectrum of the Crab nebula pulsar recorded by Romani *et al.*¹ using a superconducting microcalorimeter. An arriving photon heats the electrons in the superconductor, which is cooled just below its superconducting transition, thereby producing measurable changes in the superconducting current. (Reproduced from ref. 9.)

Optical detectors containing STJs, such as those reported by Peacock *et al.*, were proposed for use on the Hubble Space Telescope and developed by a group at the European Space Agency (ESA) (the proposed instrument was not in fact selected for the Hubble Space Telescope, but the STJs are still promising for future space missions). Now, Perryman *et al.*² have announced the ESA group's first astronomical observation with a 6×6 STJ detector array mounted on the William Herschel Telescope at La Palma in the Canary Islands. Considering that each one of the 36 detector elements requires an amplifier and a pulse-height analyser, this was no small feat. They chose the pulsar at the centre of the Crab nebula as their first target. The Crab pulsar spins 30 times a second and is one of the few pulsars that is known to emit optical pulses. Such a rapidly spinning source requires an optical detector with microsecond time resolution to distinguish pulses only 33.5 ms apart. Perryman *et al.* recorded a light curve for the pulsar over the wavelength range 310–610 nm, based on data acquired over a ten-minute interval, with an arrival-time accuracy of 5 μ s.

Similar results can be obtained using a cryogenic superconducting microcalorimeter⁵, which measures the heat produced by incident photons. Romani *et al.*¹ have measured light curves for the Crab pulsar from near-infrared to near-ultraviolet using a microcalorimeter based on a transition-edge superconducting (TES) sensor (Fig. 1). This sensor consists of tiny 18- μ m squares of 40-nm-thick tungsten, similar in size to the individual pixels in semiconductor charge-

coupled device (CCD) arrays. The device is cooled to just below the 80 mK superconducting transition of the detector elements, and a voltage is applied to the superconductor so that a current flows. The electrical resistance of the superconductor is still not quite zero, so the current will heat it up a little, quickly reaching a stable equilibrium because the hotter the electrons, the less current can flow. Now, when a photon hits the tungsten and is absorbed by an electron, the temperature of the tungsten-electron system rises a tiny amount, less current flows through the superconductor, and the change can be measured. These TES detectors are very sensitive, partly because very small current changes can be measured using a superconducting amplifier known as a SQUID (superconducting quantum interference device).

Now that superconducting detectors have been shown to work, what next? The ability to measure single photons could be extended to much longer wavelengths. The TES microcalorimeters have an energy resolution of 0.15 eV and are able to detect individual photons with wavelengths up to 4 μ m before running into false alarms from electronic noise. To get much better resolution the detector would have to be much smaller or colder, both of which are possible. Small microcalorimeters could be coupled to superconducting 'bow-tie' antennas, enabling the detection of wavelengths longer than the size of the thermometer. For even longer wavelengths, single-photon counting may not be necessary, as the sky is bright with emissions from interplanetary and interstel-

lar dust. In this case, detector sensitivity need only be better than fluctuations from random photons, a performance that is nearly feasible with today's detectors. At the other end of the spectrum the challenge is to detect extremely short wavelengths. X-ray detectors using semiconductor thermistors have already been built for rockets and satellites, and more sensitive versions using TES or STJ technology may fly on NASA's Constellation-X or ESA's XEUS missions⁶.

The STJ detector may also be pushed further. The invention of the radio frequency single electron transistor (RF-SET) by Schoelkopf *et al.*⁷ opens the door to counting single low-energy photons. Schoelkopf and colleagues' amplifier uses two STJs in series to sense the voltage on a gate capacitor. It is similar to the SQUID amplifier, which uses two STJs in parallel to measure the current in an input inductor. Both the RF-SET and SQUID are quantum-mechanical amplifiers with supreme sensitivity. Work is already underway at Yale University and NASA's Goddard Space Flight Center to test these devices for measuring currents in STJ detectors for far-infrared wavelengths of 100 μ m or longer. Another collaboration is building a 200-pixel camera using STJs for the Palomar Observatory at Caltech.

Building the circuitry to operate large arrays of superconducting detectors will require a new level of engineering. CCD detector arrays are already commonplace in digital cameras and digital video recorders, and infrared detector arrays using silicon, germanium and various alloys have also been built. The infrared devices usually have a separate amplifier for each pixel, as well as transistors to apply voltage and reset the accumulated charges. Nothing like this has yet been done for superconducting electronics, but enough principles are known for there to be high hopes. The challenge is to harness new technologies — such as the high-speed SQUID multiplexer⁸, in which multiple signals can be simultaneously transmitted — to instrument detector arrays. With broadband detection from every pixel in such an instrument, the potential benefits to astronomy are huge. ■

John C. Mather is in the Infrared Astrophysics Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

e-mail: john.mather@gsfc.nasa.gov

- Romani, R. W., Miller, A. J., Cabrera, B., Figueroa-Feliciano, F. & Nam, S. W. *Astrophys. J.* **521**, L153–L156 (1999).
- Perryman, M. A. C. *et al. Astron. Astrophys.* **346**, L30–L32 (1999).
- Peacock, A. *et al. Nature* **381**, 135–137 (1996).
- Paresce, F. *Nature News & Views* **381**, 115 (1996).
- Cabrera, B. *et al. Appl. Phys. Lett.* **73**, 735–737 (1998).
- Stahle, C. K., McCammon, D. & Irwin, K. *Phys. Today* **52**, 32–37 (1999).
- Schoelkopf, R. J., Wahlgren, P., Kozhevnikov, A. A., Delsing, P. & Prober, D. E. *Science* **280**, 1238–1242 (1998).
- Chervakov, J. A., Irwin, K. D., Grossman, E. N., Martinis, J. M. & Reintsema, C. D. *Appl. Phys. Lett.* **74**, 4043–4045 (1999).
- <http://astro.stanford.edu/users/rwr/home.html>

Controlling the cellular brakes

Peter Carmeliet

How does a cell decide to become a nerve cell and not, say, an endothelial cell? Its fate lies with certain 'master' genes, which switch on cellular programmes to differentiate precursor cells into, for example, nerve, muscle or heart tissue. Differentiation is thought to be kept in check by molecular brakes, and, on page 670 of this issue, Lyden *et al.*¹ discuss the importance of these brakes for correct development of both the neuronal system (neurogenesis) and new blood vessels (angiogenesis). The authors also report that endothelial cells switch on these differentiation brakes during tumour angiogenesis. So, by manipulating them, perhaps angiogenesis — and, hence, tumour formation — can be blocked.

Initial insight into how a cell's fate is set emerged from observations² that the basic helix-loop-helix (bHLH) transcription factor MyoD can single-handedly convert fibroblasts into muscle cells. Other bHLH proteins have since been implicated in the lineage development of blood cells (haematopoiesis), neurons, cardiac muscle (cardiogenesis) and skeletal muscle (myogenesis). They fall into two classes: the class A bHLH proteins are ubiquitously expressed, whereas the class B proteins occur in a more

tissue-restricted pattern. The HLH region mediates homo- or hetero-dimerization between class A and B proteins, and the basic domain is responsible for specific DNA binding and for transcriptionally activating genes involved in differentiation^{2,3} (Fig. 1a).

Considering the complexity of cell-fate determination, it is no surprise that bHLH proteins need to be counter-regulated. In 1990, Benezra *et al.*⁴ identified the 'inhibitor of differentiation' (Id) proteins as potential cellular brakes to prevent premature cell differentiation. Subsequent studies revealed that Id proteins can affect cell fate in a diverse manner. Although they lack the basic DNA-binding domain, Id proteins can form 'non-functional' heterodimers with bHLH factors — that is, they act as dominant-negative regulators of bHLH proteins³ (Fig. 1b). Class A bHLH proteins tend to associate strongly with Id proteins, whereas the class B proteins bind poorly. However, Id proteins may affect class B bHLH proteins by sequestering their class A partners.

Proper tissue development involves an initial phase, during which the cell precursors divide, followed by differentiation of committed cell types. The Id proteins prevent a premature transition from the growth

to differentiation phases, and Lyden *et al.*¹ now reveal that such brakes are important for neurogenesis. Using a gene-inactivation technique they knocked out expression of the Id1 and Id3 proteins from neuronal precursor cells. The result was premature withdrawal of these cells from the cell cycle, upregulation of cell-cycle inhibitors, and accelerated expression of neuronal 'differentiation effector' genes and of post-mitotic differentiation markers. The authors found that, owing to this improperly timed differentiation, there was abnormal development of the neuronal cells (Fig. 1c,d).

So what about endothelial cells? Endothelial cells in different vascular beds can acquire distinct — even opposite — differentiation characteristics. None of the transcriptional regulators thus far implicated in endothelial differentiation seems to act as a master switch of cell fate, and none of these regulators contains a bHLH motif. Nonetheless, bHLH proteins are suspected to have a hand in the fate of endothelial cells, and, indeed, Lyden *et al.* show that Id1 and Id3 negatively regulate vascular development. Most endothelial cells seem to express Id1, Id2 and Id3, a redundancy that may explain why, in most embryonic regions, blood vessels can develop in the absence of Id1 and Id3. In contrast, endothelial cells in the developing brain express only Id1 and Id3, explaining why the authors found abnormal vascular morphogenesis there.

Which endothelial bHLH transcription factors do the Id proteins inhibit? We do not yet know. Few bHLH proteins have been identified in endothelial cells, but examples include stem-cell leukaemia/T-cell acute leukaemia-1 (ref. 5), the bHLH-zip transcription factor Tfeb (ref. 6), bHLH-EC2 factor⁷ and the hypoxia-inducible bHLH-PAS transcription factors HIF-1 α and HIF-2 α (ref. 8). Several of these factors are also thought to be involved in angiogenesis, because embryos or tumours that lack them show vascular defects^{5,6,8,9}.

Does the loss of Id1 and Id3 induce premature withdrawal of endothelial cells from the cell cycle, as in neuronal cells? The abnormal vascular morphogenesis observed by Lyden *et al.* might be due to premature differentiation of the endothelial cells, which could prevent these cells from completing normal patterning and assembling into a mature vascular network. One sign of such premature differentiation could be the observed reduction in expression of vascular endothelial growth factor-2 in the Id1/Id3-deficient endothelial cells — the receptor for this growth factor is usually upregulated in proliferative endothelial cells.

Lyden and colleagues' data also touch on the question of how adult endothelial cells, after remaining quiescent for years, can suddenly become highly proliferative when angiogenesis is reinitiated during pathologi-

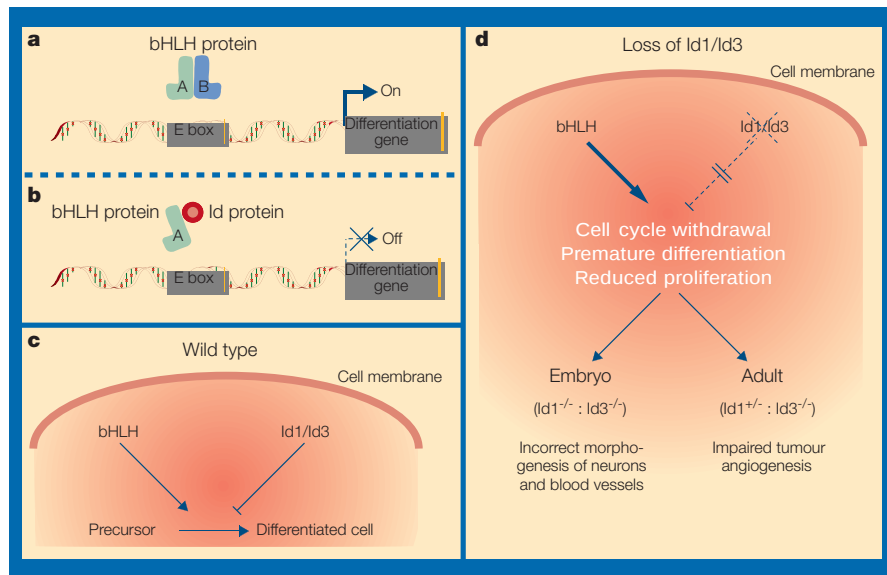


Figure 1 Relationships between the basic helix-loop-helix (bHLH) and 'inhibitor of differentiation' (Id) proteins, and cell-fate determination. a, b, Heterodimers of bHLH class A and B proteins bind to E-boxes in the promoters of differentiation genes. The Id proteins form non-functional heterodimers that cannot bind to the E-box and so prevent gene transcription. c, d, In wild-type cells, 'positive' bHLH and 'negative' Id proteins antagonistically control the differentiation of precursor cells. Lyden *et al.*¹ have found that, in the absence of Id1 and Id3, uncontrolled bHLH activity induces withdrawal from the cell cycle and premature differentiation. This leads to abnormal morphogenesis of neurons and endothelial cells in the embryo, and impaired pathological angiogenesis in the adult.

cal conditions such as in tumours. Quiescent adult endothelial cells express minimal levels of the Id proteins, whereas Id expression is upregulated in angiogenic endothelial cells. Increased levels of the Id proteins in turn upregulate expression of the $\alpha v\beta 3$ integrin — a typical marker of activated angiogenic endothelial cells. This integrin binds the matrix metalloproteinase type-2 protein, which is required for breakdown of the endothelial basement membrane (a prerequisite for migration of endothelial cells to form new blood vessels). This means that Id proteins may be involved in the switch from a quiescent to an angiogenic endothelial cell. Indeed, partial loss of Id1 and Id3 impairs tumour angiogenesis because endothelial cells fail to express the molecular tools needed to form new blood vessels.

It is tempting to speculate that Id1/Id3-deficient endothelial cells do not undergo this angiogenic switch because the lack of these differentiation inhibitors forces them into a state of (terminal?) differentiation. Tumours secrete pro-angiogenic factors, but maybe even these cannot rescue the endothelial cells. In that case, strategies to inhibit endothelial Id proteins could be used to block angiogenesis therapeutically. Furthermore, because of the differences in Id

expression between quiescent and tumour-associated endothelial cells, only the tumour should be affected if the Id proteins are inhibited — that is, this could be a safe, selective strategy. And perhaps the Id proteins could be therapeutic targets to inhibit angiogenesis in other disease states, such as retinal ischaemia or inflammation. Although much more work is needed, Lyden and colleagues have, at least, provided a promising outlook. ■

Peter Carmeliet is at the Center for Transgene Technology and Gene Therapy, Flanders Interuniversity Institute for Biotechnology, Campus Gasthuisberg, Herestraat 49, University of Leuven, Leuven B-3000, Belgium.

e-mail: peter.carmeliet@med.kuleuven.ac.be

1. Lyden, D. *et al.* *Nature* **401**, 670–677 (1999).
2. Molkentin, J. D. & Olson, E. N. *Curr. Opin. Genet. Dev.* **6**, 445–453 (1996).
3. Norton, J. D., Deed, R. W., Craggs, G. & Sablitzky, F. *Trends Cell Biol.* **8**, 58–65 (1998).
4. Benezra, R., Davis, R. L., Lookshin, D., Turner, D. L. & Weintraub, H. *Cell* **61**, 49–59 (1990).
5. Visvader, J. E., Fujiwara, Y. & Orkin, S. H. *Genes Dev.* **12**, 473–479 (1998).
6. Steingrimsson, E., Tessarollo, L., Reid, S. W., Jenkins, N. A. & Copeland, N. G. *Development* **125**, 4607–4616 (1998).
7. Quertermous, E. E., Hidai, H., Blonar, M. A. & Quertermous, T. *Proc. Natl Acad. Sci. USA* **91**, 7066–7070 (1994).
8. Semenza, G. L. *Curr. Opin. Genet. Dev.* **8**, 588–594 (1998).
9. Carmeliet, P. *et al.* *Nature* **394**, 485–490 (1998).

Device physics

Memories are made of ...

Angus Kingon

For all the amazing advances that have been made in semiconductor technology, most notably in miniaturization and processor speed, some goals remain elusive. Among those goals is realizing the 'ideal' nonvolatile memory — memory that retains information when power is removed from the device, thereby precluding the need for battery back-up. On page 682 of this issue¹, Park *et al.* describe how bismuth titanate incorporating lanthanum might be a good material for this purpose.

The problem being tackled here is that the common dynamic random access memory (DRAM), found in most computers, requires information stored in every bit to be rewritten several times per second. Some forms of nonvolatile memories are in current service (for example, for storing executable code in a processor), and they include so-called 'EEPROM' and 'flash' types which are based upon standard silicon processing. However, both suffer from disadvantages such as long write times, a limited number of write cycles and large size.

Ferroelectric random access memories (FERAMs) are promising nonvolatile alternatives. These devices use the switchable

spontaneous polarization (effectively resulting in stored charge) which is the characteristic of a ferroelectric material as the means of storing information. But the ferroelectric materials concerned are complex, and the trail leading to integration with silicon devices and commercialization has proved to be surprisingly arduous.

Although many ferroelectrics are known (see, for example, ref. 2), few have been seriously scrutinized for use in nonvolatile memories. The two main targets for development have been the PbZrO_3 – PbTiO_3 solid solution (PZT), and a related material based on $\text{SrBi}_2\text{Ta}_2\text{O}_9$ (SBT). Although PZT is in commercial — albeit limited — production at the moment, both material types have disadvantages which have limited their evolution into mainstream memory products. As Park *et al.*¹ point out, these disadvantages include the high process temperatures of SBT, which makes the material difficult to process in conjunction with silicon devices, and the polarization 'fatigue' of PZT on common platinum electrodes. This fatigue is the reduction of the polarization or stored charge with repeated switching (that is, reading and writing).

So Park and co-workers have tried a different approach. They have modified another ferroelectric material, bismuth titanate (BTO), by the solid-solution addition of lanthanum oxide, to form $\text{Bi}_{3.25}\text{La}_{0.75}\text{Ti}_3\text{O}_{12}$ (BLT) films. Films of this composition can be made at a lower temperature than SBT, meaning that they can be integrated with silicon devices more readily. Furthermore, they have a higher remanent polarization than SBT, while at the same time displaying an apparent resistance to polarization fatigue. These results are encouraging. At the least, they serve as a reminder that there are a host of possible materials yet to be considered in pursuing FERAMs. Furthermore, it is quite possible that BLT will turn out to have both the desired properties and ease of processing that will make it a superior choice to PZT and SBT.

For FERAM in general to become the dominant type of nonvolatile memory, however, it will need to be made competitive with 'flash' memories in the semiconductor mainstream rather than simply in niche markets where the competition primarily comes from EEPROM. For this, it will have to be extended from low-density memories (in the 1-megabit range) to far higher densities such as 64 megabits, and beyond. At these densities, the ferroelectric capacitor of the memory cell must shrink to sub-micrometre dimensions. In addition, adoption of the most likely cell geometries will limit the thicknesses to less than 30 nanometres.

Although it has been shown that there is no inherent size effect which limits ferroelectric switching at these dimensions, there are several obstacles to integration at such densities. Among the questions that have to be resolved are the following. First, how susceptible are these materials to hydrogen degradation? The films are exposed to hydrogen during the later stages of device processing, and the bismuth-based ferroelectrics appear to be even more susceptible to degradation after hydrogen exposure than the lead-based analogues.

Second, what is the structure of the lanthanum-substituted BLT material — is it similar to that of PZT, or is it a layered structure such as SBT and bismuth titanate? It should be noted that the phenomenon of ferroelectricity is strongly rooted in the anisotropy of the structure. The crystal structure also determines the allowed polarization directions. Furthermore, at a microstructural level, the optimum orientation of the grains within the films is determined experimentally; for SBT the approach is to make the grains randomly oriented. This may present a problem as one scales the devices for very high densities. For example, Gruverman³ has pointed out that the random orientation of SBT may result in increased bit-to-bit variability as the capacitor size is decreased.

Finally, the anisotropy and allowed polarization directions influence the switching behaviour. Switching in particular directions can result in significant local strain. In practice, this 'ferroelastic' effect manifests itself as regions of the film which do not switch. The phenomenon is not important for the low-density memories in use now, but is expected to be critical for the high-density devices. It will be interesting to study this aspect of the BLT films produced by Park *et al.*¹.

An interesting feature of the paper by Park *et al.* is their speculation regarding the origin of the fatigue-free behaviour, and their postulate that the presence of bismuth in a particular lattice site strongly affects fatigue. It is difficult to comment on their postulate, as the precise mechanisms of fatigue are not fully known, but the advent of a new material with unique properties and behaviour should lead to new insights into the mechanisms that underlie fatigue and degradation.

Are there any other materials in the offing? It is worth noting that the class of ferroelectrics known as 'relaxors' have been investigated for a number of years, but only a

handful of studies have considered their potential as nonvolatile memories. However, there is some evidence that they may be useful. For example, the work of Maria *et al.*⁴ indicates that films of the relaxor lead magnesium niobate-lead titanate (PMN-PT; $\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3 \cdot x\text{PbTiO}_3$) have attractive ferroelectric properties. There are also indications that these and related relaxors can be fatigue-free (C. A. Randall, Pennsylvania State Univ., unpublished results).

The hopes for ferroelectric memories are great, but the road there has so far been long and winding. Where it will end, we don't know, but the results provided by Park *et al.* are a step along that road. ■

Angus Kingon is in the Department of Materials Science and Engineering, North Carolina State University, Raleigh, North Carolina 27695-7919, USA.

e-mail: Angus_Kingon@ncsu.edu

1. Park, B. H., Kang, B. S., Bu, S. D., Noh, T. W., Lee, J. & Jo, W. *Nature* **401**, 682–684 (1999).
2. Mitsui, T. & Nomura, S. (eds) *Ferroelectrics and Related Substances: Oxides* Landolt-Börnstein New Series Group III, Vol. 16a (Springer, Berlin, 1981).
3. Gruverman, A. *Appl. Phys. Lett.* **75**, 1452–1454 (1999).
4. Maria, J. P., Hackenberger, W. & Trolier-McKinstry, S. *J. Appl. Phys.* **84**, 5147–5154 (1998).

Immunology

Dual personality of memory T cells

Charles R. Mackay

Often, when a pathogen first infects the body, nothing happens — at least, not for several days. But when the same pathogen is encountered a second time, the response is rapid and vigorous. This phenomenon, known as immunological memory, is the basis for vaccination, yet very little is understood about how it protects the entire body so effectively. On page 708 of this issue, Sallusto *et al.*¹ report that a newly identified cell-surface marker — the chemokine receptor CCR7 — can distinguish two subsets of T cells that carry immunological memory. These two subsets apparently circulate around the body by different pathways, and mediate the memory response in different ways. The results provide a new insight into the memory response, one that is relevant to basic immunology, as well as to vaccination strategies and the development of immunosuppressive therapies.

The immune system responds slowly to newly encountered pathogens owing to the relative lack of antigen-specific cells. New pathogens must first be transported to nearby lymphoid tissue, where the primary immune response develops and where naive lymphocytes and antigen-presenting cells interact. Two clonally expanded populations emanate from this primary response:

'effector' cells, which combat spread of the pathogen; and 'memory' cells, which guard against subsequent infections. Effector and memory cells are thought to be distributed to all tissues in the body, particularly epithelial surfaces (such as the skin and gut) where pathogens are likely to be re-encountered, so lymphocyte migration facilitates systemic immunological memory.

How can we tell the difference between naive, memory and effector T cells? This has been an elusive goal. In humans, isoforms of CD45 (a phosphatase involved in cell signalling) seemed able to distinguish naive and memory T cells. Any cell expressing the CD45RO isoform was regarded as a memory T cell, based on the fact that CD45RO⁺ (memory) T cells — but not CD45RA⁺ (naive) T cells — can respond vigorously *in vitro* to previously encountered antigens. Moreover, whereas CD45RA⁺ T cells migrate almost exclusively through lymphoid tissue², CD45RO⁺ T cells migrate throughout the body, including epithelial surfaces. The rationale is that a naive T cell has such a low probability of seeing its specific antigen that it needs to travel through the lymph nodes, which are designed for massive migration of lymphocytes and antigen sampling². In contrast, memory or

effector cells migrate mainly to peripheral tissues, providing protection at sites vulnerable to challenge by pathogens. But the problem with this system was that both memory and effector cells express CD45RO, so there was still no way to distinguish between the two.

Sallusto *et al.*¹ have now used a monoclonal antibody that recognizes the chemokine receptor CCR7 to discriminate between the two subsets of CD45RO⁺ T cells. The CCR7[−] subset has many characteristics of effector cells. For example, these cells rapidly produce effector cytokines such as interferon- γ , interleukin-4 and interleukin-5, or they express perforin granules (a feature of cytotoxic cells). Cells in the CCR7⁺ subset, on the other hand, behave more like true memory cells — they lack effector function, but the authors found that these cells could differentiate to CCR7[−] effector cells after being stimulated with antigen. So, expression of CCR7, in conjunction with CD45RO, distinguishes memory cells from effector cells.

Many cells that migrate to the lymph nodes, such as naive T cells and antigen-presenting cells, express CCR7. Moreover, the chemokines that bind to CCR7 (MIP-3 β and secondary lymphoid chemokine; SLC) are both constitutively expressed in lymph nodes. For example, SLC is produced by specialized blood vessels in the lymph nodes and provides the entry cue for CCR7⁺ cells circulating in the blood³. And the lymphocytes in mice that lack CCR7 or SLC migrate poorly through the lymph nodes^{4,5}. But CCR7 is only half the story, because cells first adhere to the walls of the lymph-node vessels using a molecule called L-selectin. So, the 'combination code' for entry of T cells to normal lymph nodes is L-selectin binding, followed by CCR7 signalling through SLC. This code is expressed by naive T cells and CCR7⁺ memory T cells (Fig. 1, overleaf), but not by certain cells that do not recirculate, such as neutrophils or monocytes. Trafficking experiments show^{2,6} that a subset of memory T cells does indeed enter lymph nodes, although we cannot yet say whether these are Sallusto and colleagues' CCR7⁺ memory T cells.

The traffic of T cells through peripheral sites differs fundamentally from that through lymph nodes. So-called 'inflammatory' chemokines and receptors are involved, together with adhesion molecules that are upregulated by inflammatory cytokines. The T cells recruited to peripheral tissues are usually very distinctive, and include CD45RO⁺ T cells expressing chemokine receptors such as CCR5, CCR2 or CCR3 (ref. 7). But Sallusto *et al.*¹ show that these T cells, which are the effector-type cells, do not express CCR7. CCR5 is a particularly good marker for a subset of effector T cells, and T cells expressing CCR5 are prominent in inflammatory dis-

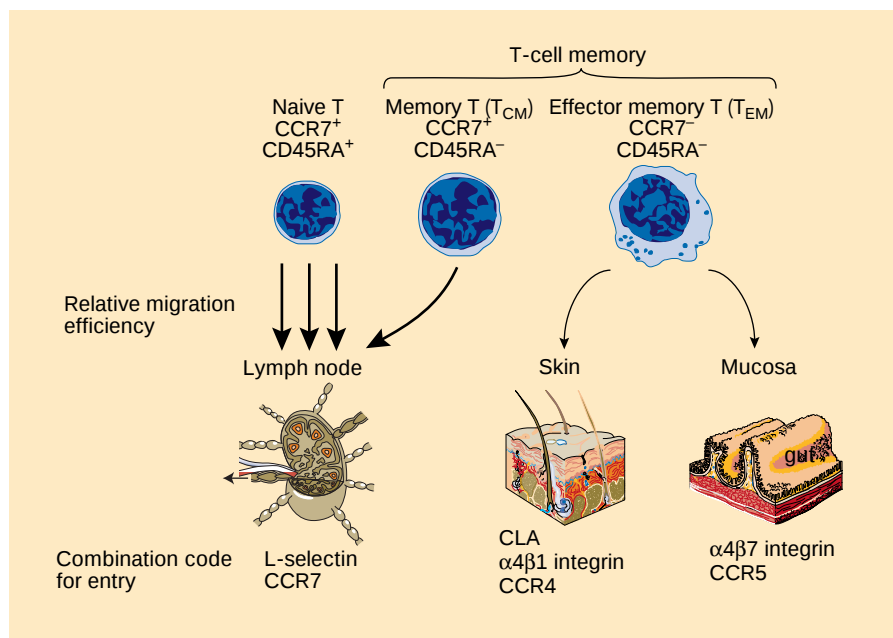


Figure 1 Two types of memory T cell with different migration preferences. Expression of CCR7 and the CD45RA isoform distinguishes three subsets of T cells: naive T cells (CCR7⁺ CD45RA⁺); a new subset, which Sallusto *et al.*¹ term central memory T cells (T_{CM}; CCR7⁺ CD45RA⁻); and effector memory T cells (T_{EM}; CCR7⁻ CD45RA⁻). Naive T cells migrate very efficiently through lymph nodes using a 'combination code' of L-selectin (an adhesion molecule) and CCR7 (a chemokine receptor). Central memory T cells probably migrate through lymph nodes, whereas effector memory T cells migrate through peripheral tissues such as the skin and mucosa. There are likely to be many combination codes for homing of T cells to the skin and mucosal tissues. CLA, cutaneous lymphocyte antigen.

ease and at mucosal surfaces⁷ (a feature that may relate to the transmission of HIV-1). In fact, it is likely that the T cells that show a degree of tissue-selective migration — preferential migration through the skin, say, rather than the gut — are effector-type T cells. For instance, the skin-homing subset of effector T cells expresses cutaneous lymphocyte-associated antigen and the chemokine receptor CCR4 (ref. 8).

The basis of immunological memory has long been contentious. According to one theory, the long-lived, recirculating memory T cells differentiate quickly to effector cells when they re-encounter an antigen. Another theory holds that constant exposure to the antigen is needed to maintain memory. Sallusto and colleagues now show that both CCR7⁺ (memory) and CCR7⁻ (effector) cells from people immunized with tetanus toxoid respond well to subsequent challenge with the toxoid *in vitro*. That is, toxoid-specific effector T cells remain alive and well, years after an initial immunization. The implication is that immunological memory is carried by both classical-type memory T cells (which Sallusto *et al.* term central memory T cells, T_{CM}) and effector T cells (now referred to as effector memory T cells, T_{EM})².

So what is the precursor-product relationship between the central memory T cell and the 'effector' memory cell? Sallusto *et al.* propose that the naive T cell differentiates

to a central memory T cell and, finally, to an effector memory cell. A more conventional idea is that effector cells are the precursors of memory cells⁹. Whatever the case, the existence of two types of memory cell with different migration pathways makes a lot of sense. Effector memory cells provide immediate front-line protection, particularly at epithelial surfaces. The central memory cells are the reserves, which can differentiate rapidly to effector memory cells in the lymph nodes in response to antigen. In conclusion, Sallusto and colleagues' study illustrates an important principle — the relationship between the functional activity of lymphocytes and their migration properties. For this reason, the chemokine receptors are proving to be excellent markers for various subsets of immune cells. ■

Charles R. Mackay is at the Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, New South Wales 2010, Australia.
e-mail: c.mackay@garvan.unsw.edu.au

1. Sallusto, F., Lenig, D., Forster, R., Lipp, M. & Lanzavecchia, A. *Nature* **401**, 708–712 (1999).
2. Mackay, C. R. *Adv. Immunol.* **53**, 217–265 (1993).
3. Gunn, M. D. *et al. Proc. Natl Acad. Sci. USA* **95**, 258–263 (1998).
4. Gunn, M. D. *et al. J. Exp. Med.* **189**, 451–460 (1999).
5. Förster, R. *et al. Cell* **99**, 23–33 (1999).
6. Williams, M. B. & Butcher, E. C. *J. Immunol.* **159**, 1746–1752 (1997).
7. Sallusto, F., Lanzavecchia, A. & Mackay, C. R. *Immunol. Today* **19**, 568–574 (1998).
8. Campbell, J. *et al. Nature* **400**, 776–780 (1999).
9. Jacob, J. & Baltimore, D. *Nature* **399**, 593–597 (1999).

Daedalus

Go with the flow

As a building material, concrete has one major defect. It cannot creep under load, even slightly; it just cracks. A building cemented with concrete mortar will develop unsightly cracks as the ground slowly settles beneath it. The old lime mortar set much more slowly, and could creep with a building as it settled in. Daedalus is now inventing a more forgiving mortar, which never truly sets at all.

In fully set concrete the load is taken by the sand grains. They are held in position by the interlocking matrix of silicate crystals formed by the setting reaction between the cement and the water. Surplus water soon evaporates, leaving a porous but unyielding solid.

So Daedalus's new concrete is designed never to dry out. DREADCO's chemists are devising a calciferous mortar loaded with hygroscopic and deliquescent chemicals which take up water avidly from the air, and 'super slurper' polymers that can absorb hundreds of times their weight of water. This novel mixture sets, but stays wet. The crystals precipitated by the setting reaction are permanently immersed in their own saturated solution.

The resulting concrete is stress-sensitive. The sand grains of the mixture still carry the compressive load. But a shear load, which might cause cracking, stresses the crystals that hold the grains in place. And just as a stressed alloy is more vulnerable to corrosion, a stressed crystal is more soluble than an unstressed one. So the stressed crystals dissolve in the solution around them, allowing the sand grains nearby to creep and relax the stress. Meanwhile, the resulting supersaturated solution precipitates new unstressed crystals which hold the grains in their new positions. The mortar has yielded safely to the creep.

DREADCO's 'Wet Cement'® will be welcomed by architects seeking to economize on foundations, and insurance companies worried by subsidence. Householders, however, may fear that permanently wet brickwork will encourage damp, mould and discoloured decor. But Wet Cement, while largely water, will have a low vapour pressure. After all, it is formulated to take up water from the air and hold it. Provided its paint or plaster covering is porous, it will remain in safe hygroscopic equilibrium. A Wet Cemented building may distort strangely under subsidence or land-heave, but will still stand firm, and look smart and house-proud.

David Jones

Familiarity breeds contempt in guppies

Male guppy fish increase their reproductive success by mating with unfamiliar females.

When President Coolidge famously attributed the sexual ardour of roosters on a government farm to a steady stream of new hens¹, he identified a strategy whereby males increase their reproductive success. By preferentially copulating with different females, a polygamous male will facilitate the spread of his genes and sire more offspring². But in order to benefit, a male must either have the cognitive skills to recognize familiar females, or behaviour patterns that increase his likelihood of encountering new ones. We show here that wild guppies, *Poecilia reticulata*, do both.

Guppies are small freshwater fish native to Trinidad and northeastern South America. They have a highly promiscuous mating system in which males engage in almost continuous courtship³. Although the mate choice of female guppies has been extensively investigated⁴, the mating preferences of males are less well understood. It is known that male guppies use phenotypic or behavioural cues that are correlated with reproductive status: for example, they prefer larger, and thus more fecund, females⁴, as well as sexually receptive ones⁵. As female guppies are monomorphic in colour pattern⁴, the identification of new partners requires a more subtle mechanism. Although these fish tend to school, forage and inspect predators in the company of familiar conspecifics⁶⁻⁸, the role of familiarity in mate choice has been largely ignored.

We examined the mating preferences of male guppies in two contexts: males confined to isolated pools or aquaria, where they repeatedly encounter the same females, and males that inhabit open rivers and so are able to search constantly for new mates. Because it takes more than 12 days for familiarity to be acquired⁹, male guppies that spend short periods of time with a group of females may treat all potential mates as novel. We therefore predicted that males confined to pools or aquaria would direct most of their courtship towards new females, whereas males from rivers would not discriminate between females from their own school and those from a different school.

Eight guppy schools, comprising 11 to 24 individuals, were captured in their entirety in Trinidad in April 1999. Four schools came from the Quare River (PS 973 797) and four from the Lower Tacarigua River (PS 792 810), neither of which had barriers to fish dispersal. The Upper Tunapuna River (PS 759 797), in contrast, forms a series of isolated pools during the dry season, in which familiarity develops, at least

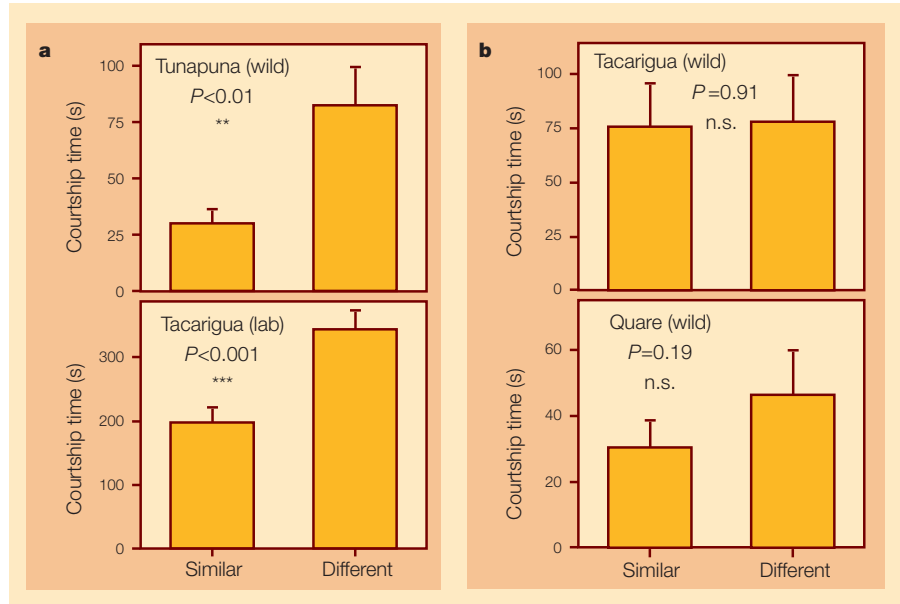


Figure 1 Males direct more courtship towards unfamiliar females. **a**, Males already familiar with females in their pool or aquarium. In both cases (Tunapuna wild, $t_{39} = 3.15$, $P = 0.003$; Tacarigua lab, $t_{30} = 4.1$, $P < 0.001$; t -tests conducted following arcsine transformation of data) there was a significant preference for unfamiliar females (shown as mean $\pm$ s.e. courtship time per fish). **b**, Males from schools in open rivers. There was no significant difference in the courtship of females from the same or a different school (Tacarigua wild, $t_{25} = 0.12$, $P = 0.91$; Quare wild, $t_{35} = 1.34$, $P = 0.19$).

among female guppies¹⁰. Ten male and ten female guppies were collected from each of four equal-sized pools towards the end of the dry season in April 1999.

All of the fish were taken to laboratory aquaria in Trinidad and allowed to acclimatize overnight with fish from the same school or pool. Tests were conducted the following day. To confirm that courtship preferences were due to familiarity, rather than being a population effect, we carried out an additional test using a laboratory stock of guppies from the Lower Tacarigua River. These fish had been held in six replicate aquaria (each containing ten males and ten females) for six weeks before the experiment.

Populations, including the laboratory population, were tested separately. The experiment was balanced such that half the males in each population first encountered females from their own group (pool, school or aquarium), whereas the other half first met females from a different group. In the trials, three similarly sized females from two different groups were allowed to settle in two identical observation tanks for one hour.

Three individually recognizable males were then introduced to one of the tanks. We observed each male for 10 min and recorded as a measure of his courtship motivation the time he spent pursuing females, displaying to them and engaging in

mating attempts^{4,5}. Males were transferred between observation tanks and their behaviour was measured again in the presence of the second set of females. This design ensured that any preferences for new females could not be attributed to differences in female exploratory behaviour¹¹.

As expected, males that were confined, and thus had become familiar with females, directed significantly more of their courtship towards unfamiliar females from different pools or tanks (Fig. 1). For fish from the open river, however, there was no significant difference in the courtship of either group (Fig. 1).

As female schools persist for extended periods in the wild⁸, males can maximize their mating opportunities by moving among them and do not need to discriminate familiar from unfamiliar females. Conversely, the ability to identify unfamiliar females offers a reproductive advantage to males that are temporarily trapped in pools. These behaviours may also result in increased gene flow. Any preferences by females for novel males^{12,13} would reinforce this outcome.

J. L. Kelley, J. A. Graves, A. E. Magurran
Institute of Environmental and Evolutionary
Biology, Bute Building,
University of St Andrews,
St Andrews, Fife KY16 9TS, UK
e-mail: jlk1@st-andrews.ac.uk

1. Wilson, J. R., Kuehn, R. E. & Beach, F. A. J. *Comp. Physiol. Psychol.* **56**, 636–644 (1963).
2. Adler, N. T. in *Biological Determinants of Sexual Behaviour* (ed. Hutchison, J. B.) 656–695 (Wiley, New York, 1978).
3. Magurran, A. E. & Seghers, B. H. *Proc. R. Soc. Lond. B* **258**, 89–92 (1994).
4. Houde, A. E. *Sex, Color, and Mate Choice in Guppies* (Princeton Univ. Press, New Jersey, 1997).
5. Liley, N. R. *Behaviour* **13** (suppl.), 1–197 (1966).
6. Milinski, M., Külling, D. & Kettler, R. *Behav. Ecol. Sociobiol.* **27**, 17–21 (1990).

Apoptosis

Searching for FLASH domains

During programmed cell death (apoptosis), a protein named FLASH is required to regulate the proteolytic cascade that ends in the death of the cell. Imai and co-workers have reported¹ that FLASH appears to be a functional analogue of two other apoptotic proteins, mammalian Apaf-1 and its nematode homologue CED-4, and that FLASH contains an amino-acid sequence motif that is homologous to the ATPase domain of Apaf-1, to the CED-4 sequence, and to a family of plant stress-resistant proteins that are apoptotic ATPases². Furthermore, FLASH contains two other domains (DRD) that are apparently related to the death-effector domain (DED)¹, an adaptor sequence that mediates interactions between proteins of the apoptosis machinery². These findings should help to explain the mechanism of action of this important protein. However, we have been unable to confirm the existence of these domains after re-examining the FLASH sequence.

We could identify no sequence similarity between FLASH and the Apaf-1/CED-4 or DED domains by searching the non-redundant protein sequence database at the NCBI using the gapped BLAST or PSI-BLAST programs^{3,4}, and over 1,000 sequences in the database were found to be more similar to the 'CED-4 homology'¹ and 'DED homology' regions of FLASH than were CED-4 or Apaf-1.

Searching databases, however, may only detect less than half of all similarities between sequences in proteins that are considered to be homologues on the basis of structural comparisons^{5,6}. Further analysis is needed, for example by direct comparison of functionally analogous proteins. We compared FLASH with Apaf-1/CED-4 and with DED-containing proteins by using the MACAW program⁷, but failed to detect any blocks of statistically significant sequence similarity (data not shown). We also used the PHI-BLAST program to assess the importance of the ATP-binding (P-loop) signature in FLASH (this program screens for similarity only those sequences that contain the specified signature), but found

7. Dugatkin, L. A. & Wilson, D. S. *Rev. Fish Biol. Fish.* **3**, 368–372 (1993).
8. Griffiths, S. W. & Magurran, A. E. *Anim. Behav.* **56**, 689–693 (1998).
9. Griffiths, S. W. & Magurran, A. E. *Anim. Behav.* **53**, 945–949 (1997).
10. Griffiths, S. W. & Magurran, A. E. *Proc. R. Soc. Lond. B* **264**, 547–551 (1997).
11. Heinrich, W. & Schröder, J. H. *Biol. Zbl.* **105**, 491–502 (1986).
12. Farr, J. A. *Evolution* **31**, 162–168 (1977).
13. Hughes, K. A. *et al. Anim. Behav.* **58**, 907–916 (1999).

no similarity to the apoptotic ATPases even in this reduced search space. The other four motifs typical of the apoptotic ATPases² are not conserved in the published alignment of FLASH with these ATPases.

To confirm the presence of DED-related domains in the FLASH sequence, we searched it with the DED profile by using the SMART system⁸ and an independent method based on the PSI-BLAST program that detects all known DED domains⁹, but we were unable to find any similarity to DED.

These tests cannot rule out a subtle similarity, although we believe that structure predictions for FLASH and phylogenetic analysis may effectively do so, at least with regard to the purported ATPase domain. Compositional complexity analysis using the SEG program¹⁰ indicates that FLASH is largely a non-globular protein (Fig. 1). The entire 'CED-4 homology' region of FLASH is predicted to be non-globular, which is incompatible with the compact structure based on a parallel β -sheet with inserted α -helices that is typical of ATPase domains¹¹. The P-loop in ATPases and GTPases is preceded by a hydrophobic β -strand, but this feature is lacking in FLASH.

The argument against structural similarity is supported by phylogenetic evidence. We have cloned and partly sequenced a human homologue of FLASH which has 67% amino-acid identity with FLASH in an alignment of 1,250 residues; the P-loop signature, however, is not conserved (data not shown; GenBank accession no. AF165161).

The structural and evolutionary evidence thus indicates that FLASH contains no globular domains with predictable functions and is not homologous to its functional

analogues. FLASH does contain a predicted coiled-coil domain (Fig. 1) which may mediate functionally important protein–protein interactions¹² and so is probably the best available lead we have from the sequence for further experiments.

Eugene V. Koonin*, L. Aravind†, Kay Hofmann‡, Jurg Tschopp§, Vishva M. Dixit¶

*NCBI, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

†Department of Biology, Texas A&M University, College Station, Texas 70843, USA

‡Bioinformatics Group, MEMOREC Stoffel GmbH, Stoeckheimer Weg 1, 50829 Köln, Germany

§Institute of Biochemistry, Lausanne Branch, University of Lausanne, Epalinges, Switzerland

¶Department of Molecular Oncology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA
e-mail: dixit@gene.com

1. Imai, Y. *et al. Nature* **398**, 777–785 (1999).
2. Aravind, L., Dixit, V. M. & Koonin, E. V. *Trends Biochem. Sci.* **24**, 47–53 (1999).
3. Altschul, S. F. *et al. Nucleic Acids Res.* **25**, 3389–3402 (1997).
4. Aravind, L. & Koonin, E. V. *J. Mol. Biol.* **287**, 1023–1040 (1999).
5. Park, J. *et al. J. Mol. Biol.* **284**, 1201–1210 (1998).
6. Brenner, S. E., Chothia, C. & Hubbard, T. J. *Proc. Natl Acad. Sci. USA* **95**, 6073–6078 (1998).
7. Schuler, G. D., Altschul, S. F. & Lipman, D. J. *Proteins* **9**, 180–190 (1991).
8. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864 (1998).
9. Chervitz, S. A. *et al. Science* **282**, 2022–2028 (1998).
10. Wootton, J. C. & Federhen, S. *Methods Enzymol.* **266**, 554–571 (1996).
11. Hubbard, T. J. P., Murzin, A. G., Brenner, S. E. & Chothia, C. *Nucleic Acids Res.* **25**, 236–239 (1997).
12. Lupas, A., Van Dyke, M. & Stock, J. *Science* **252**, 1162–1164 (1991).

Kimura, Imai and Yonehara reply — We do not consider that our inferences about FLASH function are misleading because they were deduced from functional analyses¹. We originally identified the death-effector domain (DED)-binding activity of the DED-recruiting domain (DRD) of FLASH before analysing the structural homology between the DED and DRD domains. Moreover, we deduced the self-association activity of FLASH through its central region from functional analysis of several deletion mutants of mouse FLASH. We note that self-association activity of

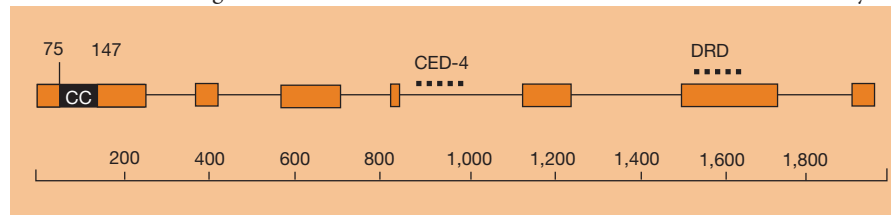


Figure 1 Diagram of the predicted domain organization of FLASH (roughly to scale). The numbered bar indicates amino-acid residue positions. Boxes, predicted globular regions; lines, predicted non-globular domains; CC, coiled-coil. Regions of alleged similarity¹ to the apoptotic ATPases (CED-4) and DED domains (DRD, or DED-related domains) are indicated by broken lines. Predicted non-globular domains were detected by using the SEG program¹⁰, with the following parameters optimized for partitioning protein sequences into globular and non-globular domains: window length, 45; trigger complexity, 3.4; extension complexity, 3.7. Coiled-coil domains were predicted using the COILS2 program¹²; boundaries of the strongly predicted coiled-coil domain are indicated.

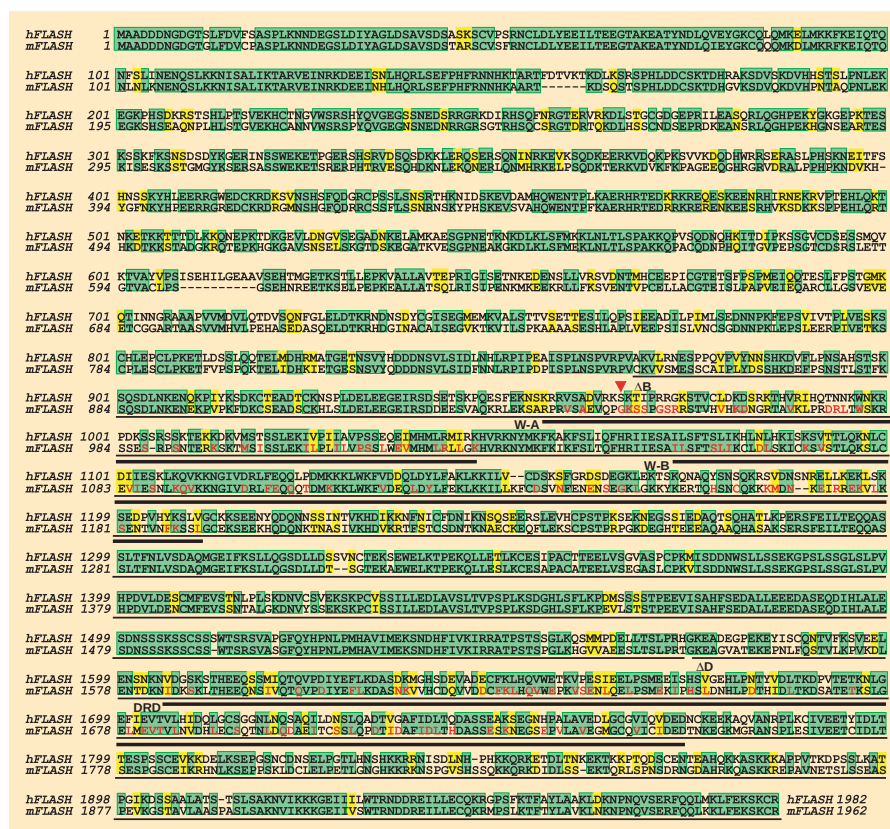


Figure 1 Comparison of deduced amino-acid sequences of mouse and human FLASH. Green and yellow boxes indicate identical and similar residues, respectively. Deletion mutants ΔB and ΔD containing 'CED-4-homologous' and DRD domains, respectively, which showed self-association and DED-recruiting activity, respectively¹, are underlined. Walker's A- and B-box-containing regions are underlined (W-A and W-B), as are the DRD domains. The amino-acid residues of mouse FLASH (red) in W-A, W-B and DRD domains are identical or similar to residues in CED-4-homologous and DED-containing proteins, as indicated previously¹. Human FLASH was cloned by using a cDNA library of human KT cells⁶ (a gift from S. Nagata). The cDNA library was screened by using fragments of human FLASH cDNA (EST clones N68740 and H50582) as probes and by standard colony hybridization procedures. The nucleotide sequence corresponding to the amino-acid sequence of human FLASH has been deposited in the GenBank database (accession no. AF154415).

CED-4-homologous proteins is required for the activation of caspases².

To determine the significance of the conserved amino-acid residues in mouse FLASH and in other apoptotic proteins, we cloned human FLASH to see whether these residues are still conserved in human FLASH (Fig. 1). Human FLASH complementary DNA showed 80% identity in the nucleotide coding sequence with mouse FLASH cDNA; in the amino-acid sequence, human FLASH shares 66% identity and 78% similarity with mouse FLASH.

In human FLASH, however, the ATP/GTP-binding motif homologous to the Walker's A-box consensus sequence³ is not conserved, with a serine residue being substituted for an essential glycine in the ATP/GTP-binding motif (arrowhead in Fig. 1). In addition, amino-acid residues that are conserved between mouse FLASH and CED-4-homologous proteins in the Walker's A-box-containing region are not conserved in human FLASH (W-A in Fig. 1). These results indicate that the ATP/GTP-binding motif of mouse FLASH is not essential and that FLASH is not a CED-4-

homologous protein. As Koonin *et al.* suggest, we could have overestimated the structural homology of FLASH with CED-4, even though these proteins are functionally homologous¹. In the self-association region (ΔB in Fig. 1), there are some well conserved domains outside the W-A region, and it needs to be determined which of these conserved domains are required for the self-association activity of FLASH.

Contrary to the inference of Koonin *et al.* (a misunderstanding that may have arisen as the result of our calling DRD a DED-related domain to indicate a DED-recruiting domain), we reported that the DRD domain of FLASH is functionally different from the DED domain¹ in that the DRD domain is unable to self-associate, whereas the DED domain can. However, there could be some conservation of amino-acid residues in the FLASH DRD domain with those in the DED domain, given that these residues are conserved between the DRD domain of mouse FLASH and the DED domain of human FLASH (Fig. 1). Moreover, Koonin *et al.* show that the DRD domain is a globular domain, so

we presume that the DRD domain has a structure that is quite similar to the DED domain.

Although the EST database has been exclusively searched using the BLAST program for DED-containing proteins, this method has never been able to identify FLASH. Having established that FLASH can bind to the DED domain¹, we cloned FLASH by using this function to detect it. Such an approach complements cloning methods based on searches of the EST database using BLAST, which has been successful for many other important apoptotic proteins, such as DED-containing FLICE and FLIP^{4,5}.

To address a more general concern regarding the specificity of our antibody, we cloned human FLASH and expressed recombinant Flag-tagged human FLASH in 293T cells. Following immunoprecipitation of human FLASH with anti-Flag antibody, our affinity-purified anti-mouse FLASH antiserum¹ was able to detect human FLASH on a western blot. Furthermore, we have shown that our antibody, which was raised against the peptide LSPNSDRNG-DAHR (from mouse FLASH), can bind to the peptide PTQDSCENTEAHQ (from human FLASH), albeit with a lower affinity than to mouse FLASH. We have also obtained monoclonal antibodies that recognize both mouse and human FLASH with associated characteristics to our affinity-purified polyclonal antibodies, and we shall report these results elsewhere.

Takaharu Kimura, Yuzuru Imai, Shin Yonehara*

*Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan
e-mail: syonehar@virus.kyoto-u.ac.jp*

1. Imai, Y. *et al.* *Nature* **398**, 777–785 (1999).
2. Walker, J. E. *et al.* *EMBO J.* **1**, 945–951 (1982).
3. Yang, X., Chang, H. Y. & Baltimore, D. *Science* **281**, 1355–1357 (1998).
4. Muzio, M. *et al.* *Cell* **85**, 817–827 (1996).
5. Irmiler, M. *et al.* *Nature* **388**, 190–195 (1997).
6. Itoh, N. *et al.* *Cell* **66**, 233–243 (1991).

Materials

Transformation of diamond to graphite

Despite almost forty years of trying, no one has managed to transform diamond into graphite under pressure¹, or find out what the pressure limit for diamond might be². If diamond were to behave like other group IV elements, such as silicon, germanium or tin, it would transform under compressive indentation to the β -tin structure³, but it does not^{2,4}. Here we use micro-Raman spectroscopy to determine what happens to diamond when it is subjected to high contact compression as a result of pressing a

sharp diamond indenter against its surface⁴. We find that, under this non-hydrostatic compression, diamond at the point of indentation is transformed into disordered graphite. This discovery may eventually lead to the more efficient machining of diamond.

In this non-destructive spectroscopy, a laser is used to analyse a tiny area of surface (one square micrometre) under the microscope. This technique has previously been used to study compression-induced phase transformations in silicon, germanium, other semiconductors and ceramics^{5,6}. When applied to diamond, in the area where the indenter touched the surface of the diamond crystal, no diamond remains after indentation (Fig. 1): disordered graphite is found instead, and the indenter tip also becomes graphitic. The presence of graphitic carbon was confirmed by Raman spectroscopy, transmission electron microscopy and electron diffraction.

The indentation tests were performed using standard Vickers pyramidal indenters under loads of 30 newtons on the (111) surface of a synthetic diamond crystal, and analysis was carried out using a Dilor LabRam microspectrometer equipped with an automated X–Y stage. An argon ion laser with a wavelength of 514.5 nm was focused onto 1 μm^2 of surface to excite the Raman effect. The nominal compressive stress at the maximum load (load divided by indentation area) was slightly below 100 gigapascals, but this can be affected by micro-cracking and phase transformations. It was calculated that pressures at low loads can reach 900 GPa^{7,8}.

Our finding that graphite is formed seems to contradict the well-known fact that graphite transforms into diamond under pressure. However, we must take into account the difference between compression and hydrostatic pressure. In the case of contact loading (indentation or scratching), the diamond structure is sheared as well as having its volume changed, not only compressing the bond lengths but also changing the angle between the bonds³. This shear strain leads to a phase transformation of diamond into a high-compression state, which is probably followed during decompression by graphite formation, as this is the thermodynamically stable phase of carbon at ambient pressure. The possibility of transformation in diamond under shear in polishing experiments^{9,10} was based on analysis of wear debris and wear tracks on diamond surfaces. Our indentation experiments indicate that the formation of a 'black layer' when diamond is polished or cut may result from compression-induced graphitization.

The transformation may occur either through a metallic phase similar to that in silicon^{3,6,11} or as a result of a loss of structural

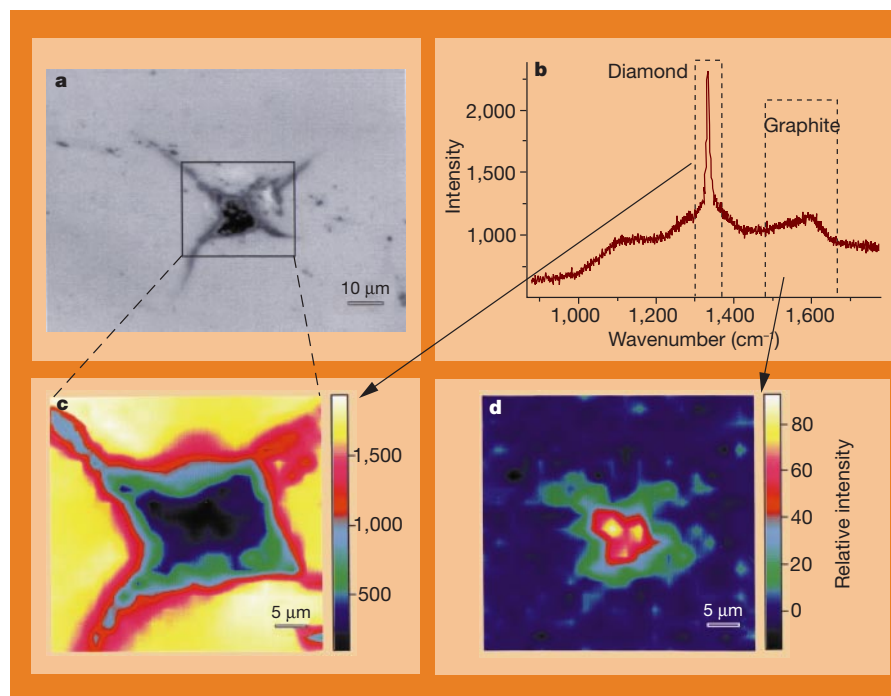


Figure 1 Raman analysis of indentation on diamond. **a**, Optical micrograph of indentation on diamond. **b**, Typical Raman spectrum of the indentation region. **c**, **d**, Raman-band intensity maps of diamond (**c**) and graphite (**d**). Decreasing colour intensity (from yellow to black) indicates a decrease in the Raman-band intensity. The black-blue area in **c** corresponds to non-diamond carbon; the bright area in **d** corresponds to graphitic carbon in the centre of the indentation.

stability during loading or unloading. Graphite forms on unloading because the diamond–graphite transformation is accompanied by an increase in volume. Atomic force microscopy reveals that the whole indentation area is slightly elevated above the crystal surface owing to expansion of the material under the indenter. This is in agreement with the low density of wear debris from polishing experiments⁹, estimated to be 1.8 g per cm^3 .

Our results demonstrate that the pressure limit for diamond can be reached by using non-hydrostatic contact loading, without recourse to superpresses or sophisticated pressure cells. This shows that diamond is not stable under high non-hydrostatic pressure and that it experiences a phase transformation for which graphitic carbon is the final transformation product after unloading. It also may explain why diamond is so hard and resistant to wear, and why diamond can be polished and machined by diamond, whereas tools that are harder than the materials to be machined are usually required.

Finally, these findings may lead to a method for machining diamond with higher efficiency¹⁰ and at a lower cost. For example, if a sharp indenter (or hundreds of indenters) is pressed against diamond, this might transform the surface layer into

graphite. Another tool, moving behind the first one, might then be able to scratch off this soft graphitic carbon, and the cycle could be repeated until the desired depth of cut is achieved. If diamond can be machined more cheaply in this way, it might then find many new engineering applications.

Yury G. Gogotsi*, **Andreas Kailer†‡**,
Klaus G. Nickel†

**Department of Mechanical Engineering,
University of Illinois, 842 West Taylor Street,
Chicago, Illinois 60607-7022, USA*

*†Institut für Mineralogie, Petrologie und Geochemie,
Universität Tübingen, Wilhelmstrasse 56,
72074 Tübingen, Germany*

*‡Present address: Fraunhofer Institute für
Werkstoffmechanik, Wöhlerstrasse 11,
79108 Freiburg, Germany*

1. Yin, M. T. & Cohen, M. L. *Phys. Rev. Lett.* **50**, 2006–2009 (1983).
2. Mao, H. K. & Hemley, R. J. *Nature* **351**, 721–724 (1991).
3. Gilman, J. J. *Science* **261**, 1436–1439 (1993).
4. Gogotsi, Y. G., Kailer, A. & Nickel, K. G. *J. Appl. Phys.* **84**, 1299–1304 (1998).
5. Gogotsi, Y. G., Kailer, A. & Nickel, K. G. *Mater. Res. Innov.* **1**, 3–9 (1997).
6. Kailer, A., Gogotsi, Y. G. & Nickel, K. G. *J. Appl. Phys.* **81**, 3057–3063 (1997).
7. Ruoff, A. L. & Luo, H. J. *J. Appl. Phys.* **70**, 2066–2070 (1991).
8. Ruoff, A. L. & Wanagel, J. *Science* **198**, 1037–1038 (1977).
9. Grillo, S. E. & Field, J. E. *J. Phys. D* **30**, 202–209 (1997).
10. Van Bouwelen, F. M., Brown, L. M. & Field, J. E. *Ind. Diamond Rev.* **57**, 21–25 (1997).
11. Cahn, R. W. *Nature* **357**, 645–646 (1992).

brieF communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com/author/guide.html#7.4) or on request from nature@nature.com

Thermodynamic control of hurricane intensity

Kerry A. Emanuel

Program in Atmospheres, Oceans, and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

To establish useful warning systems for hurricanes, it is necessary to accurately predict both hurricane intensity and track. But although the forecasting of hurricane tracks has improved over the past 30 years, the factors that control the intensity of hurricanes are still poorly understood, leading to almost no reliability in forecasts of hurricane intensity evolution. Efforts to improve intensity forecasts have focused almost exclusively on characterizing the dynamical interactions between hurricanes and their atmospheric environment. Here I use a simple numerical model to demonstrate that, in most cases, the evolution of hurricane intensity depends mainly on three factors: the storm's initial intensity, the thermodynamic state of the atmosphere through which it moves, and the heat exchange with the upper layer of the ocean under the core of the hurricane. Such a limited number of controlling factors offers hope that, given an accurate forecast of a hurricane's track, its intensity can be reliably forecast using very simple models.

Forecasts of the tracks of hurricanes have improved steadily over the past three decades¹, owing to a combination of better observations and much improved numerical models. These improvements, coupled with advances in warning systems and preparedness for emergencies, have brought about a significant decline in loss of life in the USA in spite of a near doubling of the coastal population during this period. At the same time, the economic vulnerability to hurricanes has increased dramatically. It has been estimated² that a repeat of the Miami hurricane of 1926 would incur \$75 billion in insured loss, compromising the entire US insurance industry. Among the many costs associated with hurricanes is the expense of evacuation. In practice, many more people are evacuated than was necessary in hindsight, owing to uncertainties in the forecast of both the track and intensity of the storm. Evacuation in the face of a marginal hurricane is usually unnecessary, but is often carried out because hurricanes can intensify rapidly and unexpectedly.

In contrast to the improvement in track forecasts, there has been comparatively little advance in predictions of intensity¹ (as measured, for example, by maximum surface wind speed), in spite of the application of sophisticated numerical models. The best intensity forecasts today are statistically based³. Most of the research literature on hurricane intensity focuses on the pre-storm sea surface temperature and certain properties of the atmospheric environment, such as the vertical shear of the horizontal wind and dynamical features such as disturbances in the upper troposphere⁴. This remains so, even though it is well known that hurricanes alter the surface temperature of the ocean over which they pass⁵ and that a mere 2.5 K decrease in ocean surface temperature near the core of the storm would suffice to shut down energy production entirely⁶. Simulations with coupled atmosphere–ocean models^{6–8} confirm that interaction with the ocean is a strong negative feedback on storm intensity. During the Atlantic hurricane season of 1998, guidance based on coupled-model simulations was provided to forecasters for the first time.

Although there is much hope that three-dimensional coupled models will lead to better understanding of the factors that control hurricane intensity and to increased reliability ('skill') of hurricane intensity forecasts, the present generation of models may not have enough horizontal resolution to capture the full intensity of extreme storms. (Fortunately, it is probably not necessary to capture full storm intensity in order to achieve a good track forecast.)

Here I show that an accurate account of the evolution of storm intensity can be achieved using a very simple coupled ocean–atmosphere model in which the atmospheric component is cast in

a transformed radial coordinate that greatly increases horizontal resolution in the critical region around the eyewall, the ring of intense convection that surrounds the eye of the storm. This is so even though the atmospheric component is axisymmetric and therefore excludes interactions with vertical wind shear and dynamical features of the atmospheric environment. This demonstrates that, once storm genesis has occurred, much of the evolution of storm intensity is controlled by its initial intensity together with the thermodynamic properties of the atmosphere and upper ocean along the storm track.

The model

The atmospheric model assumes that the storm is axisymmetric, and that the airflow is never very far from a state in which the horizontal and vertical pressure gradient accelerations are balanced by centrifugal and gravitational accelerations, respectively. It also assumes that the vortex is always close to a state of neutral stability to a combination of gravitational and centrifugal convection ('slantwise convection'). These constraints place very strong restrictions on the structure of the vortex so that, with the exception of the water-vapour distribution, the vertical structure is determined by a very limited set of variables. Moist convection is represented by one-dimensional plumes whose mass flux is determined in such a way as to ensure approximate entropy equilibrium of the boundary layer. The model variables are cast in 'potential radius' coordinates¹⁰. Potential radius (R) is proportional to the square root of the absolute angular momentum per unit mass about the storm centre and is defined by $fR^2 = 2rV + fr^2$, where V is the velocity of air flowing around the storm, r is the physical radius and f is the Coriolis parameter, which is twice the local vertical component of the Earth's angular velocity.

In the runs presented here, there are 50 nodes that span 1,000 km, giving an average resolution of 20 km; however, the resolution is substantially finer than this in regions of high vorticity, such as the eyewall. A complete description of the model is given elsewhere¹¹. When run with a fixed sea surface temperature and a fixed atmospheric environmental temperature profile, and provided that the vortex specified at the start of the model integration is strong enough, the model vortex amplifies over a period of 4–5 days right up to its potential intensity (see upper curve in Fig. 1a). The potential intensity is the maximum steady intensity a storm can achieve based on its energy cycle, in which the heat input by evaporation from the ocean, multiplied by a thermodynamic efficiency, is balanced by mechanical dissipation in the storm's

atmospheric boundary layer¹². It is given by

$$V^2 = \frac{C_k}{C_D} \frac{T_s - T_o}{T_o} (k_s - k_a) \quad (1)$$

where V is the maximum wind speed, C_k and C_D are dimensionless exchange coefficients for enthalpy and momentum, T_s and T_o are the absolute temperatures of the sea surface and storm top, and k_s and k_a are the specific enthalpies of the air at saturation at the ocean surface and ambient boundary layer air, respectively. (That the outflow rather than inflow temperature appears in the denominator of equation (1) is due to the fact that the dissipative heating in the storm's boundary layer recycles some of what would otherwise be waste heat back into the thermodynamic cycle of the storm¹³.)

Whereas model storms usually spin up to their potential intensity and remain at that intensity indefinitely, real hurricanes seldom behave that way; in fact most hurricanes experience a sharp decline shortly after achieving their peak intensity¹⁴. Moreover, very few real storms ever achieve the potential intensity given by equation (1). From records of previous storms together with the climatology of potential intensity, there is a nearly uniform probability that a given

hurricane will achieve any intensity between marginal hurricane force and the potential intensity¹⁴.

The axisymmetric hurricane model is coupled to a one-dimensional ocean model in a unique way. First, it is assumed that the hurricane responds principally to sea surface temperature changes under its eyewall, and that these can be closely approximated by sea surface temperature changes under that part of the eyewall that lies along the storm track. Second, the evolution of sea surface temperature along the storm track up until the time that the centre of the storm arrives can be approximated as arising entirely from one-dimensional stirring of each vertical column, with no influence from its neighbours. (The horizontal exchange of enthalpy between oceanic columns is ignored.) Finally, the mixing is approximated by assuming that a bulk Richardson number, relating the velocity of the ocean's mixed layer to the jump in temperature across the base of that layer, remains constant¹⁵ as the ocean mixed layer is accelerated by the wind stress imposed from the passing storm. Thus the ocean model consists merely of a set of one-dimensional ocean columns along the storm track, whose temperature is changed only through vertical mixing in each column. The temperature stratification

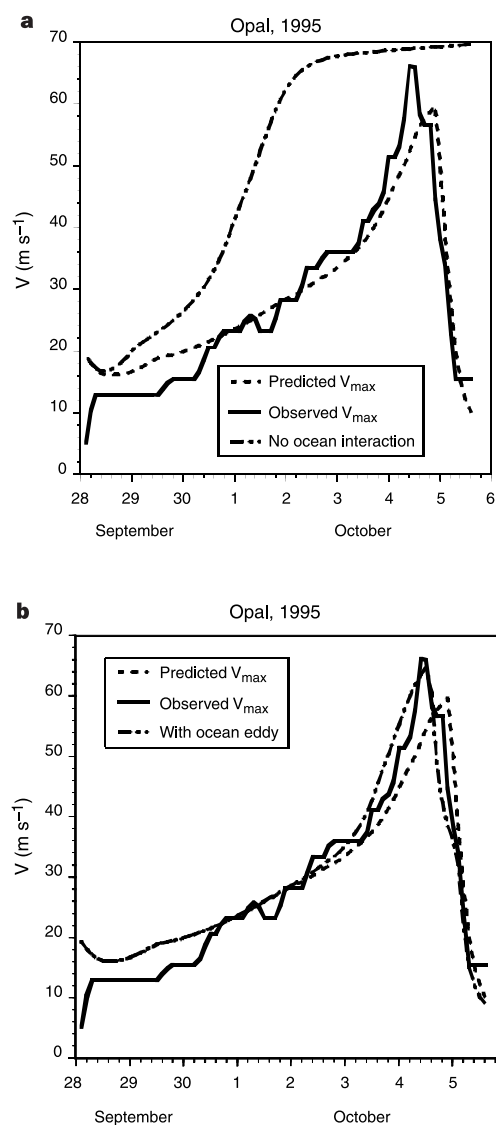


Figure 1 Evolution of the maximum wind speed in Hurricane Opal. In **a**, the solid line shows the observed evolution, the dashed line shows the modelled evolution, and the dash-dot line shows evolution modelled without ocean interaction. In **b**, the dash-dot line shows evolution modelled in the presence of a warm ocean eddy.

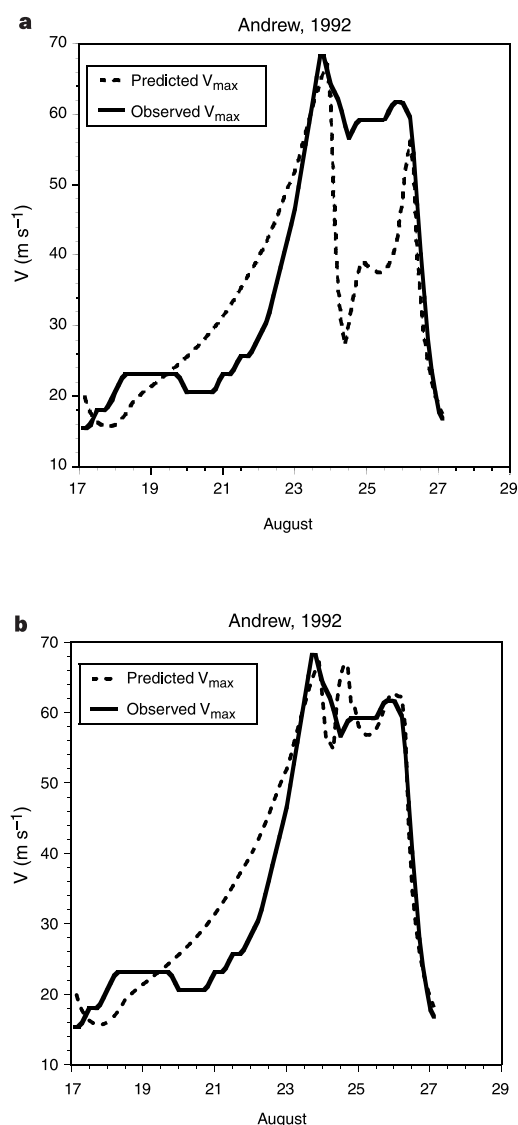


Figure 2 Evolution of the maximum wind speed in Hurricane Andrew. Solid lines are observations and dashed lines are modelled values. The nominal model run is shown in **a**, while the run in **b** takes into account the swamp in southern Florida and the shallow continental shelf to its west (see text for details).

below the mixed layer is set to a constant in the runs described here. This very simple formulation has been shown to lead to a storm intensity evolution that is virtually indistinguishable from that of the same hurricane model coupled to a three-dimensional ocean model, in the case of a steadily moving storm¹⁶.

For each event, the model is initialized using a synthetic warm-core vortex. In each of the cases discussed below, the geometry of the vortex is identical, though in principle it can be varied according to the size of the real system. The maximum wind speed of this initial vortex is matched to the observed wind speed at the beginning of the initial period of intensification of the observed system. Also, the initial degree of saturation of the inner storm core is specified so as to achieve the observed, initial rate of intensification. These are crucial steps, as the subsequent evolution is quite sensitive to the initial state. Apart from this initial matching of the model and observed maximum wind speeds, no adjustment of the model vortex towards observations is made.

Two properties of the storm's environment along its observed track are specified from monthly mean climatology: the potential maximum wind speed¹⁷ and the ocean mixed-layer depth¹⁸. The former is interpolated from the 2.5° grid on which it was supplied to the observed storm position; the latter was similarly interpolated from a 1° grid to the observed storm position. Both data sets were

also linearly interpolated to the actual date, assigning the monthly mean climatology to the 15th day of each month. These monthly climatologies were formed using many years of data; no year-to-year variations are accounted for. A third data set, on a 1° grid, was also used to specify ocean depths along the observed storm track. This was used to detect landfall, and also to reveal those situations when the ocean mixed layer extends right to the ocean floor, so that surface cooling by mixing cannot occur. The landfall algorithm is one of maximum simplicity: when the centre of the storm passes over land, the coefficient of surface enthalpy flux is set to zero everywhere. Although this is unrealistic, in practice the strongest effects are under the eyewall, whose passage over land occurs nearly at the same time that the storm centre makes landfall. (The small differences in timing are comparable to the six-hour temporal resolution of the observational data.)

In each of the cases presented below, the evolution of maximum surface wind speed in the model is compared to the observed evolution; no attempt has been made to compare the evolutions of model and observed storm structure, as the latter is not available in any convenient form. It should be borne in mind that not all of the reported wind speeds are directly measured by aircraft or radar; some are partially subjective estimates based on satellite imagery. (Here again, the readily available data archive does not document

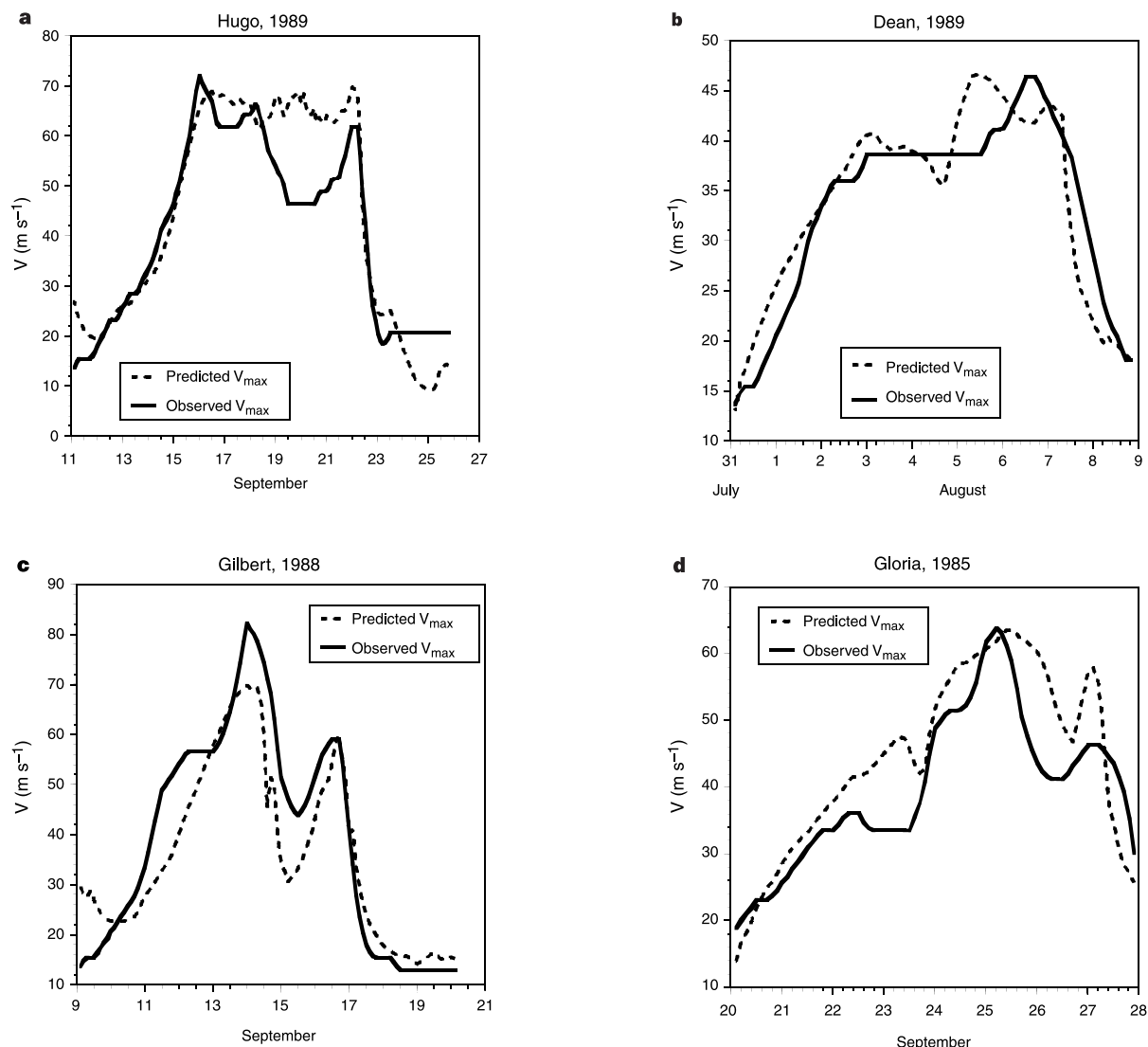


Figure 3 Evolution of maximum wind speed in several hurricanes. **a**, Hurricane Hugo; **b**, Hurricane Dean; **c**, Hurricane Gilbert; and **d**, Hurricane Gloria. In all cases the predicted

and the observed maximum wind speeds are shown by dashed and solid lines, respectively.

the source of the data, but it is safe to assume that most reported wind speeds in storms within one day of landfall in the USA are based on reliable *in situ* or radar measurements within six hours of the reporting time.)

Model skill

Some well simulated examples. Figure 1 shows the observed and modelled evolution of Hurricane Opal, which moved through the Gulf of Mexico in October, 1995, making landfall in northwestern Florida. Also shown in Fig. 1a (but not in subsequent figures) is the evolution that would have occurred had the ocean temperature remained fixed in time, demonstrating the crucial role of ocean interaction. Previous studies of this event have emphasized the role of an approaching upper-tropospheric disturbance⁴ or the observed presence of a warm ocean eddy at about the time of maximum intensification, but Fig. 1a shows that most of the evolution can be accounted for without these effects (I. Ginis, personal communication). The main influence of the approaching upper-tropospheric disturbance was to accelerate the forward motion of the storm, thereby decreasing the ocean cooling. Insertion into the model of a warm ocean eddy of about the dimensions and magnitude of that observed did result in a small but noticeable increase in the peak intensity of the storm, as shown in Fig. 1b.

Hurricane Andrew developed east of the Bahamas in August, 1992, and then moved westward across the southern tip of Florida, into the Gulf of Mexico, and then northwestward, making landfall again in Louisiana. It was the most expensive natural disaster in US history, incurring more than \$28 billion in damage. The evolution of Hurricane Andrew's intensity is shown in Fig. 2a. Here the modelled evolution departs noticeably from the observed in several respects. In its early stages, the model storm intensifies while the observed storm in fact weakens. Operational forecasters at the time attributed this weakening to the presence of substantial vertical wind shear, an effect not accounted for in this model. More spectacularly, the model intensity declines far more rapidly than observed after making landfall in southern Florida. Two important model deficiencies may come into play here: first, the southern tip of Florida is not dry land but rather a swamp, so that the assumption of vanishing surface heat flux may be extreme. Second, the resolution of the ocean depth data set was not high enough to account accurately for the presence of a shallow shelf extending westward from the southern tip of Florida. In reality, the ocean mixed layer over which Hurricane Andrew moved probably extended right to

the sea floor for the first ten hours or so after the storm left Florida. In Fig. 2b, the model has been modified to account for the actual depth of the sea floor along the storm track, and the surface enthalpy exchange coefficient has been reduced by only one-half while the storm is over southern Florida. This illustrates how very sensitive hurricane intensity is to the nature of the underlying surface.

Hurricane Hugo moved through the northern Caribbean and then up over the Sargasso Sea, making landfall in South Carolina in September, 1989. Figure 3a compares the actual and modelled storm evolution. The simulation is quite good, except when the storm is over the Sargasso Sea, in which case the model overestimates its actual intensity. There is considerable evidence that Hugo was affected by vertical wind shear during this time.

Hurricane Dean moved westward over the tropical North Atlantic to just north of the Virgin Islands, then turned north, moving over open waters until it struck southeastern Newfoundland in early August, 1989. It never exceeded marginal hurricane intensity. Figure 3b compares the predicted and modelled intensities of that storm.

Hurricane Gilbert, in September 1988, was the most intense hurricane ever recorded in the Atlantic region. It moved westward over the Caribbean Sea, striking Jamaica and the Yucatan peninsula before passing into the Gulf of Mexico. Figure 3c shows that whereas

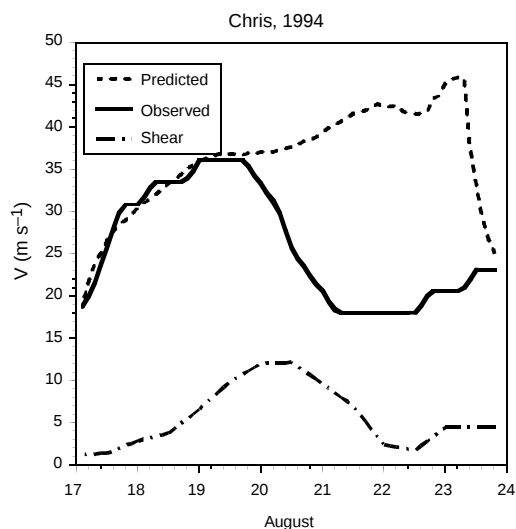


Figure 4 Evolution of the maximum wind speed in Hurricane Chris. The solid line shows observations, and the dashed lines are modelled values; the dash-dot line shows an estimate of the magnitude of the environmental vertical wind shear at the storm centre.

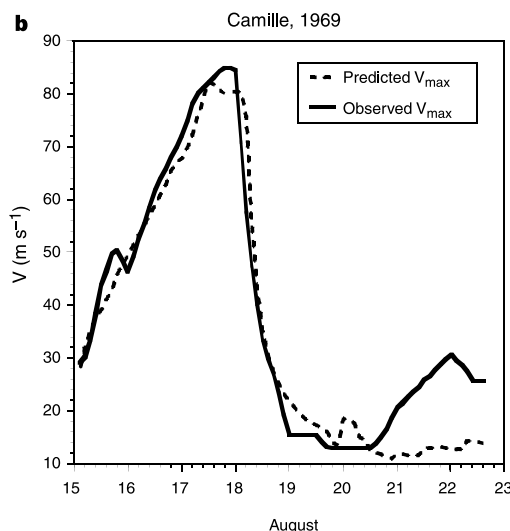
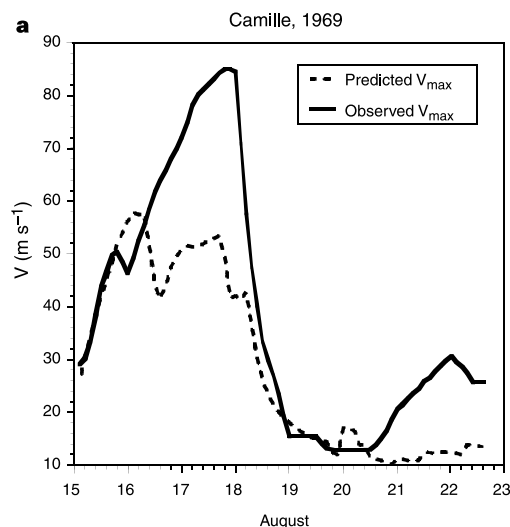


Figure 5 Evolution of the maximum wind speed in Hurricane Camille. Solid lines are observations, and dashed lines are modelled values. The nominal model run is shown in **a**, while in **b** the ocean interaction is omitted.

the character of the storm's intensity evolution was well simulated, the peak intensity was substantially underestimated.

Hurricane Gloria formed off Africa, and affected the US east coast in late September, 1985. Figure 3d shows that the evolution of its intensity was well simulated in most respects.

Some notable failures. Especially in their early stages of development, tropical cyclones are susceptible to suppression through the effects of vertical wind shear, or enhancement by dynamical interactions with other weather systems in the high troposphere. These effects have been emphasized in many previous investigations. We have found several examples in which the skill of the present model is significantly compromised by such effects. An example of the first effect is demonstrated by Fig. 4, which shows the model prediction of the wind-speed evolution of marginal hurricane Chris in 1994, together with an estimate of the magnitude of the observed environmental vertical wind shear at the storm centre (J. Kaplan, personal communication). This hurricane appears to have been severely limited by the shear, which reaches its peak intensity just as the hurricane goes into decline. We have also found several cases in which the present model underpredicts the intensity of events that were apparently affected by dynamical interactions with their environment. We emphasize, however, that such effects are significant in only a small fraction of the cases that we have examined.

Figure 5a shows the predicted and observed evolutions of Hurricane Camille of 1969, the only category 5 hurricane to strike the US mainland in the 100-year period of record, making landfall near Biloxi, Mississippi. The intensity of the storm is very badly underpredicted. Figure 5b shows that if the sea surface temperature is held fixed, the simulation is quite good. One of the interesting features of the Gulf of Mexico is the Loop Current, an extension of the Gulf Stream that loops up towards the coast of Alabama and Mississippi but whose exact position is somewhat variable. No manifestation of this current was evident in the climatological ocean data interpolated to the observed positions of Camille. There is some evidence that Camille moved right along the axis of the Loop Current, which has a locally deep, warm mixed layer¹⁹. It has been suggested²⁰ that most of the very severe hurricanes that affect the Gulf coast move along the Loop Current.

Discussion

The simulations presented here suggest that, once tropical cyclones reach tropical-storm strength, their intensity evolution is controlled mostly by their initial intensity together with the thermodynamic profile of the atmosphere and upper ocean through which they move. Factors such as vertical wind shear and dynamical interactions with the environment, emphasized in previous work, appear to be strongly influential mostly during the formative stages, when the storms are comparatively weak. Storm intensity is particularly sensitive to the thermodynamic structure of the upper ocean, and it

is evident that in at least some cases (for example, Hurricane Camille) the climatological specification of the ocean thermal structure is insufficient. Accurate prediction of hurricane intensity in these cases probably requires accurate measurement of the upper-ocean thermal structure ahead of the storm. The simulations presented here offer hope that even with a very simple model, hurricane intensity can be predicted with useful skill as far in advance as an accurate track prediction can be made. Such track predictions require three-dimensional models able to account for the full range of interactions between the storm and its environment. □

Received 6 May; accepted 26 August 1999.

- DeMaria, M. & Kaplan, J. An operational evaluation of a statistical hurricane intensity prediction scheme (SHIPS). Preprints of the 22nd Conf. on Hurricanes and Tropical Meteorology 280–281 (American Meteorological Soc., Boston, 1997).
- Landsea, C. W. & Pielke, R. A. Jr. Normalized hurricane damages in the United States: 1925–95. *Weath. Forecast.* **13**, 621–631 (1998).
- DeMaria, M. & Kaplan, J. A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Weath. Forecast.* **9**, 209–220 (1994).
- Bosart, L. F., Bracken, W. E. & Molinari, J. Environmental influences on the rapid intensification of Hurricane Opal (1995) over the Gulf of Mexico. Preprints of the 23rd Conf. on Hurricanes and Tropical Meteorology 983–984 (American Meteorological Soc., Boston, 1999).
- Price, J. F. Upper ocean response to a hurricane. *J. Phys. Oceanogr.* **11**, 153–175 (1981).
- Gallacher, P. C., Rotunno, R. & Emanuel, K. A. Tropical cyclogenesis in a coupled ocean-atmosphere model. Preprints of the 18th Conf. on Hurricanes and Tropical Meteorology (American Meteorological Soc., Boston, 1989).
- Schade, L. R. & Emanuel, K. A. The ocean's effect on the intensity of tropical cyclones: Results from a simple coupled atmosphere-ocean model. *J. Atmos. Sci.* **56**, 642–651 (1999).
- Khain, A. & Ginis, I. The mutual response of a moving tropical cyclone and the ocean. *Beitrage Phys. Atmos.* **64**, 125–141 (1991).
- Emanuel, K. A. The Lagrangian parcel dynamics of moist symmetric instability. *J. Atmos. Sci.* **40**, 2368–2376 (1983).
- Schubert, W. H. & Hack, J. J. Transformed Eliassen-balanced vortex model. *J. Atmos. Sci.* **40**, 1571–1583 (1983).
- Emanuel, K. A. The behavior of a simple hurricane model using a convective scheme based on subcloud-layer entropy equilibrium. *J. Atmos. Sci.* **52**, 3959–3968 (1995).
- Emanuel, K. A. An air-sea interaction theory for tropical cyclones. Part I. *J. Atmos. Sci.* **42**, 1062–1071 (1986).
- Bister, M. & Emanuel, K. A. Dissipative heating and hurricane intensity. *Meteorol. Atmos. Phys.* **65**, 233–240 (1998).
- Emanuel, K. A. A statistical analysis of tropical cyclone intensity. *Mon. Weath. Rev.* **127**, (in the press).
- Price, J. F. Internal wake of a moving storm, Part I: Scales, energy budget, and observations. *J. Phys. Oceanogr.* **13**, 949–965 (1983).
- Schade, L. R. A physical interpretation of SST-feedback. Preprints of the 22nd Conf. on Hurricanes and Tropical Meteorology 439–440 (American Meteorological Soc., Boston, 1997).
- Bister, M. & Emanuel, K. Hurricane climatological potential intensity maps and tables. (cited 01/05/99) (<http://www-paoc.mit.edu/~emanuel/pcmin/climo.html>)/22/03/97)
- Levitus, S. *Climatological Atlas of the World Ocean* (NOAA Prof. Pap. No. 13, US Gov. Printing Office, Washington, DC, 1982).
- Ly, L. N. & Kantha, L. H. A numerical study of the nonlinear interaction of Hurricane Camille with the Gulf of Mexico loop current. *Oceanol. Acta* **16**, 341–348 (1993).
- Cooper, C. in *Proc. Offshore Tech. Conf.* 213–222 (Offshore Tech. Conf., Richardson, Texas, 1992).

Acknowledgements

The vertical wind shear data were supplied by Dr. J. Kaplan of the NOAA/ERL Hurricane Research Division. I thank H. Willoughby for comments on the manuscript.

Correspondence and requests for materials should be addressed to K.A.E. (e-mail: emanuel@texmex.mit.edu).

Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts

David Lyden^{*,†}, Alison Z. Young^{†,‡}, David Zagzag[§], Wei Yan[‡], William Gerald^{||}, Richard O'Reilly^{*}, Bernhard L. Bader[¶], Richard O. Hynes[#], Yuan Zhuang[☆], Katia Manova^{**} & Robert Benezra[‡]

Departments of [‡] Cell Biology, ^{**} Molecular Cytology Core Facility, ^{||} Pathology and ^{*} Pediatrics, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

[§] Department of Pathology, Division of Neuropathology, Kaplan Cancer Center, New York, New York 10016, USA

[#] Howard Hughes Medical Institute for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[¶] Max-Planck-Institute for Biochemistry, Department of Proteinchemistry, D-82152 Martinsried near Munich, Germany

[☆] Department of Immunology, Duke University Medical Center, Durham, North Carolina 27710, USA

[†] These authors contributed equally to this work

Id proteins may control cell differentiation by interfering with DNA binding of transcription factors. Here we show that targeted disruption of the dominant negative helix–loop–helix proteins Id1 and Id3 in mice results in premature withdrawal of neuroblasts from the cell cycle and expression of neural-specific differentiation markers. The Id1–Id3 double knockout mice also display vascular malformations in the forebrain and an absence of branching and sprouting of blood vessels into the neuroectoderm. As angiogenesis both in the brain and in tumours requires invasion of avascular tissue by endothelial cells, we examined the Id knockout mice for their ability to support the growth of tumour xenografts. Three different tumours failed to grow and/or metastasize in Id1^{−/−}Id3^{−/−} mice, and any tumour growth present showed poor vascularization and extensive necrosis. Thus, the Id genes are required to maintain the timing of neuronal differentiation in the embryo and invasiveness of the vasculature. Because the Id genes are expressed at very low levels in adults, they make attractive new targets for anti-angiogenic drug design.

The Id proteins are a family of four related proteins implicated in the control of differentiation and cell-cycle progression in organisms ranging from flies to man (for a review see ref. 1). Id proteins exert this control by preventing transcription factors from binding DNA by direct physical interaction. Primary targets for Id proteins are the basic helix–loop–helix (bHLH) transcription factors which regulate cell-type-specific gene expression and the expression of cell-cycle regulatory genes (for a review see ref. 2). As Id lacks a basic DNA-binding domain, heterodimers between Id and bHLH proteins cannot bind DNA. This dominant negative mode of regulation is also used by members of the leucine zipper and homeodomain transcription factor families^{3,4}.

Biochemical and genetic data have established that Id acts primarily by sequestering the ubiquitously expressed bHLH proteins known as E proteins (products of the E2A, E2-2 and HEB genes), preventing them from binding DNA alone or as heterodimers with tissue-restricted bHLH proteins^{1,5}. As predicted, overexpression of Id blocks bHLH-mediated transcription in a wide variety of cell types¹. Because bHLH complexes regulated by Id may have varied effects on the differentiated state of a cell, it is not surprising that overexpression or upregulation of Id has been associated with a block to differentiation¹, the onset of differentiation^{6–9} or an alteration in cell fate^{10–12} in different cell contexts.

We addressed whether Id proteins are required to control differentiation in mammalian systems. Targeted disruption of Id1, Id2 or Id3 in mice yields the expected number of live births, but mice lacking either Id2 or Id3 show haematopoietic abnormalities^{13,14}. As all Id proteins behave similarly biochemically, we sought to inactivate two members of the family with overlapping expression patterns, Id1 and Id3, to define further their role in mammalian development.

Id1 and Id3 are co-expressed temporally and spatially during murine neurogenesis^{15–17} and angiogenesis¹⁷, two processes described here as being affected in Id1–Id3 double knockout

animals. In general terms, Id1, Id2 and Id3 are expressed in dividing neuroblasts of the central nervous system (CNS) up to about embryonic day (E) 12.5, after which Id2 expression persists in presumptive neurons that are undergoing maturation in both the future cerebellum and cerebral cortex^{8,17}. In adult animals, Id1 and Id3 are no longer expressed in the brain, but Id2 expression remains in the Purkinje cells of the cerebellum (ref. 8, and this study).

Angiogenesis, the branching and sprouting of capillaries from pre-existing blood vessels, occurs in the yolk sac and in the embryo, particularly in the brain (for a review see ref. 18). Signalling from both VEGF and Tie-2 receptors has been implicated in this process, as well as in tumour angiogenesis; however, little is known of the involvement of bHLH proteins during these processes. Id1 and Id3, but not Id2, are expressed in endothelial cells in the brain, whereas Id1, Id2 and Id3 are expressed in endothelial cells throughout the rest of the embryo during development (ref. 17, and this study).

We show that Id1 or Id3 expression is required for maintaining the timing of neuronal differentiation and proper angiogenesis in the neuroectoderm during mouse development. Importantly, partial reduction of Id dosage also results in an angiogenic defect in adult mice which blocks the vascularization, growth and metastasis of tumour xenografts.

Id1^{−/−}Id3^{−/−} mice are inviable

Mice lacking one to four Id1 and Id3 alleles were generated by intercrossing Id1^{+/−}Id3^{+/−} mice. Offspring lacking 1–3 Id alleles in any combination were indistinguishable from the wild type, but no animals lacking all four Id alleles were born. To determine when Id1^{−/−}Id3^{−/−} mice die, Id1^{−/−}Id3^{+/−} mice were intercrossed and the embryos examined from E8.5 to birth. Between E8.5 and E10.5, Id1^{−/−}Id3^{+/−}, Id1^{−/−}Id3^{+/−} and Id1^{−/−}Id3^{−/−} embryos were represented in a 1:2:1 mendelian ratio. Id1^{−/−}Id3^{−/−} embryos were grossly normal up to E10.5 but reduced in

size by 30% at E11.5 and E12.5 (Fig. 1a, b). By E12.5, the mutants exhibited cranial haemorrhage (Fig. 1b) and no embryos survived beyond E13.5, indicating that expression of either *Id1* or *Id3* is essential for viability.

The ganglionic eminences of *Id1*^{-/-}*Id3*^{-/-} embryos developed cavitation lesions (Fig. 1c, d). Areas of hypocellularity formed at E11.5, and by E12.5 coalesced into a cavity. Aberrant capillaries flanked the cavity. They ruptured at E13.5, resulting in haemorrhage throughout the ventricular system (Fig. 1d).

Id inhibits neural differentiation

The smaller brain size in *Id1*^{-/-}*Id3*^{-/-} mutants may be due to a decrease in proliferation of neuroblasts or an increase in apoptosis. At E10.5, no significant differences in apoptosis were observed between wild-type and *Id1*^{-/-}*Id3*^{-/-} embryos (data not shown). Similarly, at E10.5, no difference was found when cell proliferation was measured by immunodetection of Ki-67, a nuclear antigen expressed in all proliferating cells except those in G0 or early G1. In E11.5 mutants, however, there were fewer proliferating cells in the neuroepithelium of the telencephalon (Fig. 1e, f) and the rhombencephalon (data not shown).

The withdrawal of neuroblasts from the cell cycle might be accompanied by altered expression of cell-cycle regulatory proteins. In *Id1*-null embryonic fibroblasts, p16, the cyclin D/cdk4 inhibitor, is upregulated (R. Alani and R.B., unpublished data). Consistent with this observation, p16 (Fig. 1i, j) and p27 (data not shown) were upregulated in the mutant neopallial cortex at E11.5. The expression patterns of post-mitotic differentiation markers, microtubule-associated protein 2 (MAP2) and unique β -tubulin (TUJ1), were complementary to that of Ki-67 (Fig. 1g, h; and data not shown).

Regulation of neural bHLH expression by Id

To understand the molecular mechanism of premature neural differentiation in the mutants, we examined the expression patterns of neural-specific bHLH genes. Like myogenic bHLH proteins, there may be a hierarchy of activities in the neural bHLH family with the determination genes (*MATH1*, *MATH3*, *MASH-1*, *Ngn1*

and *Ngn2*) inducing the expression of differentiation effectors (*NeuroD1*, *NeuroD2* and *MATH2*)¹⁹.

NeuroD1 expression was similar in wild-type and mutant embryos at E9.5 and E10.5, but was increased in the ganglionic eminences at E11.5 (Fig. 2a, b). No differences were observed in *NeuroD2* and *NeuroD3* expression. In E11.5 mutants, *MATH1* expression was more extensive in the rhombencephalon (Fig. 2h), *MATH2* was enhanced and more extensive in the ganglionic eminences (Fig. 2d) and dorsal rhombencephalon (Fig. 2j), and *MATH3* was more extensive in the dorsal telencephalon and ganglionic eminences (Fig. 2f). By E12.5, no differences were detectable for these markers. In the ventral telencephalon, the areas of premature *MATH3* expression were more restricted than those of *MATH2* and *NeuroD1* (Fig. 2b, d, f).

No changes in the expression of *BF-1* (ref. 20), *Dix-2* (ref. 21), *Emx-2* (ref. 22), *Pax-6* (ref. 23) and *Shh* (ref. 24) were observed (data not shown), indicating that premature neurogenesis in the mutants is not the result of alterations in telencephalic patterning. These observations, the overlapping expression of *Id1* and *Id3* in the neuroepithelium¹⁷, and the lack of a neural phenotype in *Id1* or *Id3* single knockout mice indicate that either *Id1* or *Id3* is required for normal neurogenesis.

Vascular defects in the brain

The aberrant endothelial cells observed in the ganglionic eminences of *Id1*^{-/-}*Id3*^{-/-} embryos express CD31 normally (Fig. 3a–d). They formed enlarged, dilated blood vessels at E11.5 (Fig. 3d) and an anastomotic network at E12.5 (Fig. 3b). Laminin (Fig. 3e, f) and fibronectin (data not shown) are expressed normally in the basement membrane. By laminin immunostaining, an absence of branching and sprouting of capillaries into the neuroectoderm of the ganglionic eminences was detected (Fig. 3e, f), demonstrating abnormal angiogenesis in the mutants. This is consistent with expression of *Id1* (Fig. 3i) and *Id3* (data not shown) in normal vasculature in the CNS during development. Although *Id1*, *Id2* and *Id3* are expressed in blood vessels outside the CNS (data not shown), *Id2* expression is absent from blood vessels in the CNS

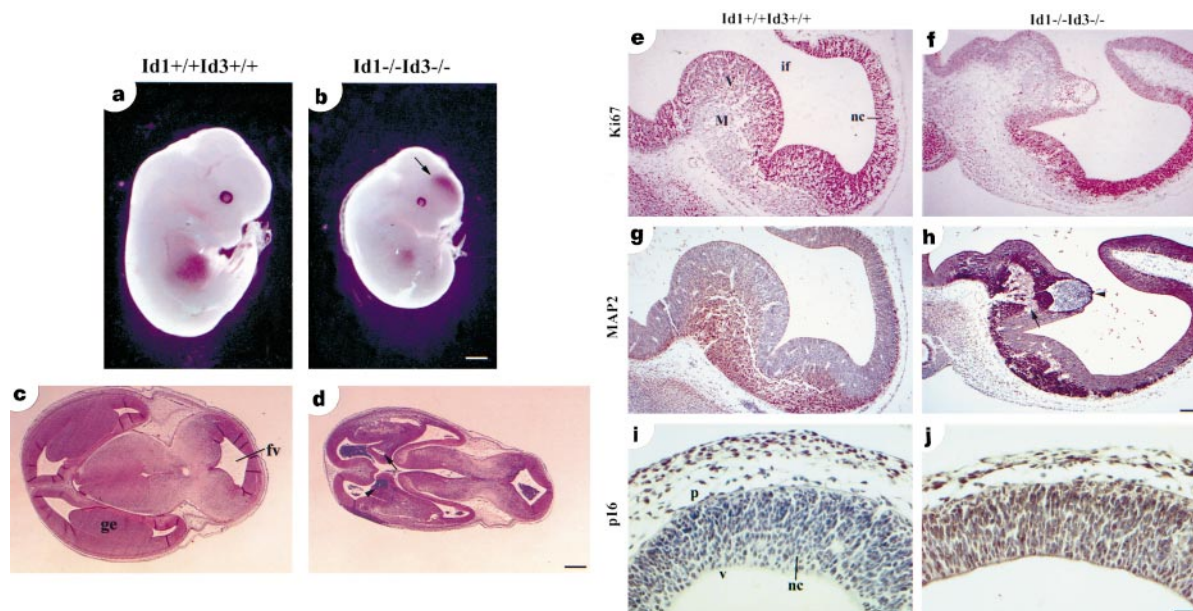


Figure 1 *Id1*^{-/-}*Id3*^{-/-} embryos exhibit cavitation lesions, cranial haemorrhage and premature neuronal differentiation. **a–d**, E12.5. **a**, Normal embryo. **b**, Mutant littermate with cranial haemorrhage (arrow). **c**, **d**, Transverse brain sections. Arrow points to cavities in the ganglionic eminences; red blood cells are seen within cavitation lesions (arrowhead) and in the ventricles. **e–j**, Immunodetection of Ki67 (**e**, **f**), MAP2 (**g**, **h**) and

p16 (**i**, **j**) in adjacent transverse sections of control and mutant E11.5 brains. Arrowhead in **h** indicates cavitation lesion, arrow points to mutant capillaries. Abbreviations: fv, fourth ventricle; ge, ganglionic eminence; V, ventricular zone; M, mantle layer; if, interventricular foramen of Monro; nc, neopallial cortex; v, ventricular surface; p, pial surface. Scale bar 400 μ m (**a**, **b**); 200 μ m (**c**, **d**); 100 μ m (**e–h**); 25 μ m (**i–j**).

(Fig. 3j).

At E12.5, reduced expression of vascular endothelial growth factor (VEGF), Flk-1 (VEGF receptor 2) and smooth muscle α -actin was first noted within the vascular malformation, but not in blood vessels outside the ganglionic eminences in the mutants (see Supplementary Information).

Vasculogenesis is normal in the mutants. The formation of the major vessels, intersomitic vessels, perineural vascular plexus and endocardium is unimpaired, and yolk sac blood islands of E8.5 mutant embryos are normal (see Supplementary Information).

The cavitation lesions in $Id1^{-/-}Id3^{-/-}$ mice resemble the defect in αv -integrin-null mice²⁵. However, no differences in the expression of $Id1$ or $Id3$ were observed in αv -integrin-null mice or vice versa (data not shown). In addition, differences in the phenotypes of these mice were noted. Capillary sprouts are present in the neuroectoderm of αv -integrin-null mice (Fig. 3g, h) unlike Id mutants, indicating that the angiogenic defect in $Id1^{-/-}Id3^{-/-}$ mice may not be an obligate consequence of hypocellularity. Also, αv -integrin-null embryos display only dilated, not malformed blood vessels.

Reduced Id dosage inhibits tumour growth

Angiogenesis both in the neuroectoderm and in tumours requires the invasion of a preformed, avascular mass by the endothelium. $Id1-4$ are not expressed in normal adult vasculature (Fig. 3k, l), but $Id1$ (data not shown) and $Id3$ are expressed in endothelial cells in high-grade human neural tumours (Fig. 3m–p). To determine whether $Id1$ and $Id3$ play a role in tumour angiogenesis, $Id1^{-/-}Id3^{-/-}$ mice and wild-type littermates were challenged with intradermal injections of various tumour cell lines.

Wild-type mice (129Sv, C57BL/6) inoculated with B6RV2 lymphoma cells have a rapid increase in tumour mass and die at 24.1 ± 7.1 days (mean $\pm$ s.d.; Fig. 4a, Table 1). In striking contrast, $Id1^{-/-}Id3^{-/-}$ mice have a small peak of tumour growth which completely regressed after 14 days (Fig. 4a). All $Id1^{-/-}Id3^{-/-}$ mice remained healthy for 540 days until they were killed (Table 1). A similar result was observed with a murine breast cancer cell line B-CA (Fig. 4b, Table 1).

In $Id1^{-/-}Id3^{+/+}$ mice, B6RV2 tumours also regressed, albeit more slowly, and the rate of B-CA tumour growth was intermediate between those of wild-type and $Id1^{-/-}Id3^{-/-}$ mice (Fig. 4a, b). Although the effect from the loss of one Id allele was surprising, such dosage effects are not uncommon in the bHLH family^{26–28}. It is unlikely that during targeted disruption of one of the Id genes a dominant gain-of-function mutation was generated at a nearby locus, as the loss of a single $Id1$ (Fig. 4) or $Id3$ allele (data not shown) produces a very similar phenotype.

Lewis lung carcinoma (LLC) cells were also used in the xenograft assay; however, these cells continued to proliferate in all strains of mutant mice, albeit at a significantly lower rate and with prolonged survival time relative to wild-type mice (Fig. 4c, Table 1). This difference may be due to the absence of metastatic lesions in the Id knockout mice (Table 1) and/or differences in the histology of the tumours (see below).

Angiogenic defects in tumours

In wild-type mice, B6RV2 tumours showed blood vessel infiltration, whereas in $Id1^{-/-}Id3^{-/-}$ animals little vascularization is observed (Fig. 5Aa, b). In agreement, CD31/PECAM staining showed normal

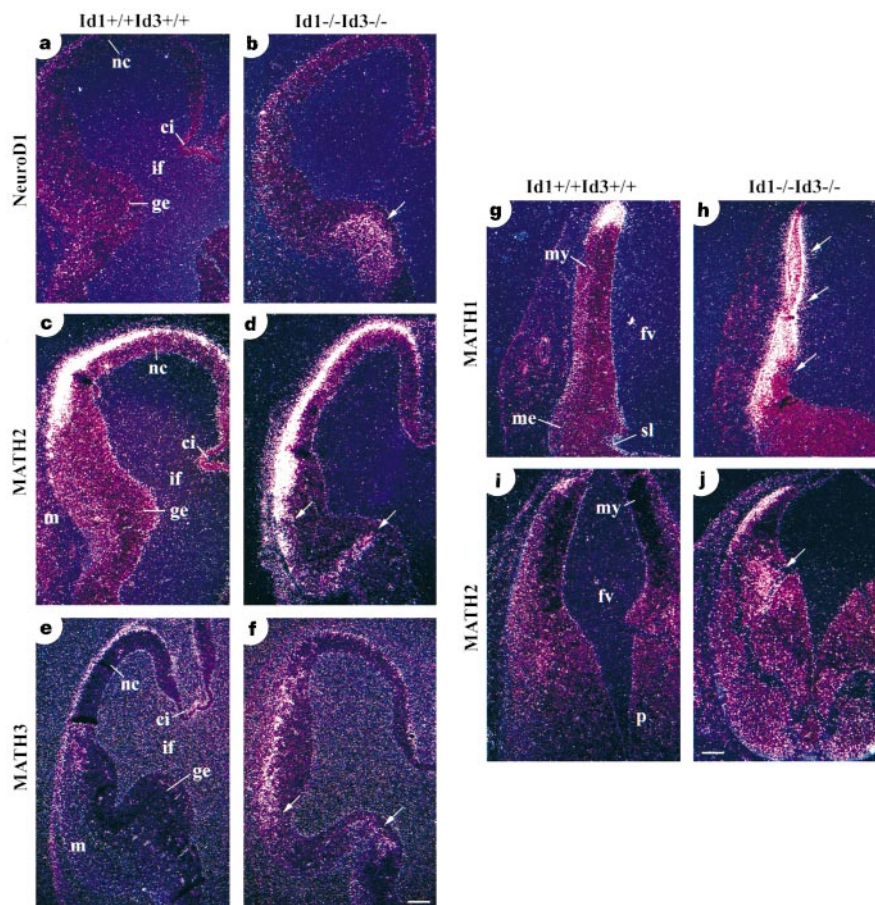


Figure 2 Premature and ectopic expression of NeuroD1, MATH1, MATH2 and MATH3 in E11.5 $Id1^{-/-}Id3^{-/-}$ embryonic brain. Arrows in **b, d, f, h, j** indicate the regions of premature or ectopic expression of these genes in mutant brains compared with wild-type

brains. Abbreviations: nc, neopallial cortex; ci, choroid invagination; if, interventricular foramen of Monro; ge, ganglionic eminence; m, marginal layer; my, myelencephalon; me, metencephalon; sl, sulcus limitans; fv, fourth ventricle; p, pons. Scale bar, 100 μ m.

Table 1 Outcome of tumour cell implantation experiments

Tumour cell type	Genotype	Overall survival (total animals)	Survival duration (days)	Metastases (total animals)
B6RV2*	Id1 ^{+/+} Id3 ^{-/-}	(22/22)	540‡	(0/22) M
	Id1 ^{+/+} Id3 ^{+/+}	(15/15)	540‡	(0/15) M
	Id1 ^{+/+} Id3 ^{+/+}	(0/20)	24.1 ± 7.1	(20/20) G [28.1 ± 8.2]
B-CA†	Id1 ^{+/+} Id3 ^{-/-}	(15/15)	330‡	(0/15) M
	Id1 ^{+/+} Id3 ^{+/+}	(0/10)	60‡	(0/10) M
	Id1 ^{+/+} Id3 ^{+/+}	(0/16)	60‡	(2/16) G [1]
LLC†	Id1 ^{+/+} Id3 ^{-/-}	(0/20)	40.2 ± 2.3	(0/20) M
	Id1 ^{+/+} Id3 ^{+/+}	(0/15)	34.2 ± 3.7	(3/15) M
	Id1 ^{+/+} Id3 ^{+/+}	(0/57)	18.9 ± 3.6	(57/57) G [7.9 ± 4.2]

G, gross; M, microscopic disease. [], total metastases per animal; mean ± s.d.

* Mesenteric lymph node metastases.

† Lung metastases.

‡ Killed.

capillaries in wild-type mice (Fig. 5Ba, c; 16.0 ± 3.2 blood vessels per 200× field) and stunted, occluded vessels in the Id knockout animals (Fig. 5Bb, d; 2.7 ± 1.4 blood vessels per 200× frame). LLC tumours were similar histologically, except that in Id1^{+/+}Id3^{-/-} mice necrotic tissue and haemorrhage predominated (Fig. 5Ac, d, Ca, b).

After LLC cell implantation, a striking difference in metastases to the lungs occurred in wild-type animals (57 out of 57 had widespread, vascularized metastasis; mean 7.9 ± 4.2 nodules per lung), compared with mutant animals (0 out of 20 Id1^{+/+}Id3^{-/-} mice had lung lesions) (Fig. 5Cc, d, Table 1). Three out of fifteen

Id1^{+/+}Id3^{+/+} animals had small, well encapsulated, avascular lesions in the lung (Fig. 5Ce, Table 1). Although the Id knockout animals eventually die, it is not due to metastases.

Mode of Id action in tumour angiogenesis

To determine whether the failure of the LLC cells to metastasize in the Id knockout animals was solely due to a failure of the cells to enter the blood stream, LLC cells were injected into the tail veins of wild-type and Id1^{+/+}Id3^{-/-} mice, and lung metastases were quantified. Wild-type animals (6 out of 6) developed extensive metastasis after 8 days (data not shown), whereas only 1 out of 6 Id1^{+/+}Id3^{-/-} animals had a single well encapsulated tumour after 21 days. Thus, the homing and/or establishment of the LLC cells is defective in the mutant animals.

We considered whether the resistance to tumour growth in the Id knockout animals was due to immune-mediated rejection; however, we observed no immune-cell infiltration in tumours in the Id knockout animals, and tumour growth was independent of the genetic background (see Figs 4 and 5). In addition, ablation of NK cells with an anti-asialo antibody in the Id knockout animals failed to restore the ability of the lymphoma cell line to grow or metastasize (data not shown). Nonetheless, to formally rule out this possibility, tumour cells were implanted into the cornea, which is avascular and has no inherent cell-mediated immunity. LLC cells showed extensive tumour growth and blood vessel infiltration in wild-type animals after 8 days (Fig. 5Da). In contrast, minimal tumour growth and patchy haemorrhages in the absence of a vascular network were observed in mutant mice (Fig. 5Db). A similar result occurred with B6RV2 cells (data not shown).

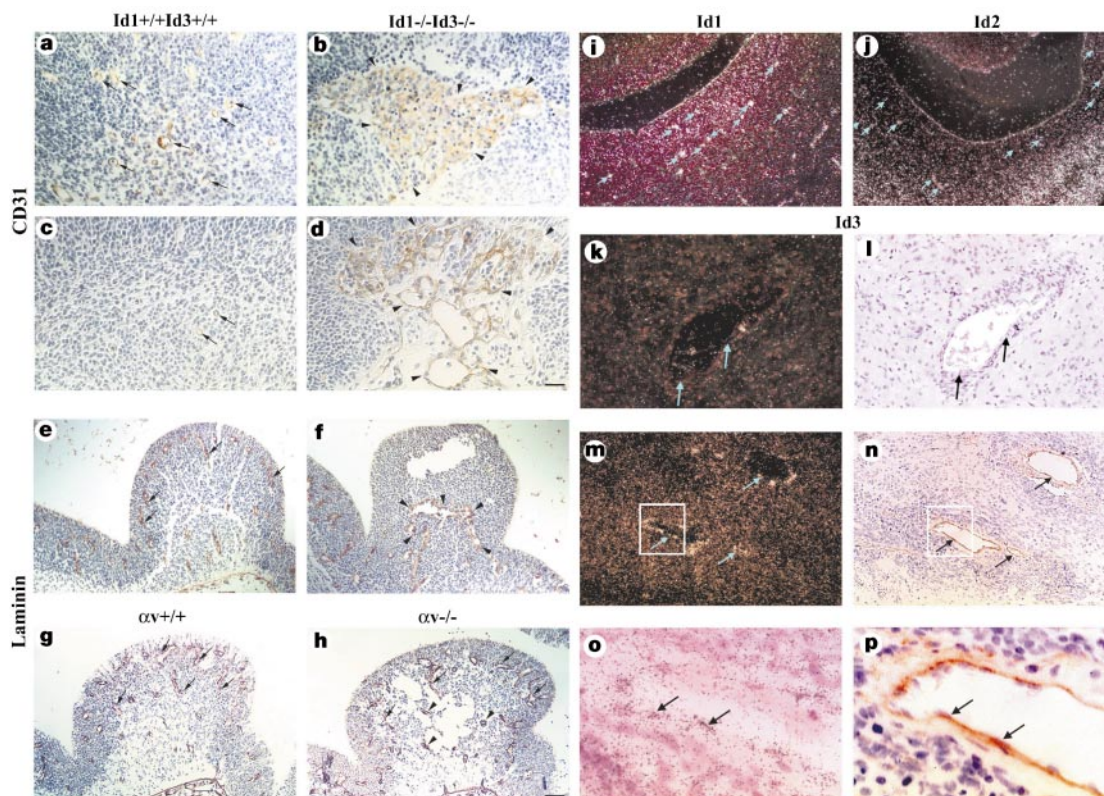


Figure 3 Defects of angiogenesis in Id1^{-/-}Id3^{-/-} embryos and Id expression in blood vessels. Blood vessels in wild-type and mutant ganglionic eminences by CD31 (a–d) and laminin immunodetection (e–h). Arrows in a, c, e, g indicate normal blood vessels, arrowheads in b, d, f, h outline mutant blood vessels. Arrows in h point to capillary sprouts in αv -integrin mutant neuroectoderm. i–p, Blood vessels in the brain, indicated by arrows. i, j, Id1 and Id2 expression in wild-type E12.5 ganglionic eminence. k, l, Id3 expression in adult cerebrum. No Id1–4 expression is found in blood vessels of brain (6

samples), adrenal gland (3), colon (3), lung (2), lymph node (3), kidney (2) and skeletal muscle (6) (data not shown). m, o, Id3 in the endothelial cells in medulloblastoma; n, p, the same vessels stained for laminin. Id1 and Id3 are expressed in blood vessels in glioblastoma multiforme (12) disseminated medulloblastoma (2) and stage III/IV neuroblastoma (14) (data not shown). Scale bar, 25 μ m (a–d, o–p); 50 μ m (e–h, k–l); 100 μ m (i, j, m, n).

The association of soluble MMP2 metalloproteinase with $\alpha v\beta 3$ -integrin has been implicated in tumour angiogenesis²⁹; therefore, we examined the levels of αv -integrin and MMP2 on the endothelial cells of the tumours. Colocalization of CD31(PECAM) and $\alpha v\beta 3$ -integrin was noted in B6RV2 tumour vessels from wild-type mice (Fig. 6Aa, c). Although CD31(PECAM) staining identified aberrant blood vessels in tumours from Id knockout mice, no αv -integrin expression was detected (Fig. 6Ab, d). As $\alpha v\beta 3$ -integrin recruits soluble MMP2 metalloproteinase²⁹, the expression of MMP2 on the surface of the tumour vasculature should also be affected. MMP2 staining was detected in tumour vessels in a wild-type host but was absent on endothelial cells in Id1^{+/+}Id3^{-/-} animals (Fig. 6Ae, f). We also noted that VEGF and Flk-1 expression was identical in the endothelium of wild-type and mutant tumours (data not shown), in contrast to that in the ganglionic eminences. This and the observation that αv -integrin or MMP2 expression is not affected in the ganglionic eminences in the Id1^{-/-}Id3^{-/-} embryos (data not shown) indicate that the effects of Id loss are probably influenced by the cellular environment of the endothelial cells.

Deficient metalloproteinase activity in tumour blood vessels should result in a thickening of the extracellular matrix (ECM). Electron microscopy of a typical capillary within the B6RV2 tumours grown in wild-type mice shows endothelial cells (E) adjacent to a relatively narrow layer of ECM (Fig. 6Ba, between arrows). In Id1^{+/+}Id3^{-/-} animals, however, the lumens of the capillaries are obstructed by cytoplasmic projections and the ECM shows marked thickening (Fig. 6Bb, between arrows).

Thus, the lack of active MMP2 on the surface of Id1^{+/+}Id3^{-/-} tumour endothelial cells could account for their failure to sprout and form a functional network. This would ultimately result in regression of the tumour mass.

Discussion

We have shown that at least one copy of the Id1 or Id3 gene is required in mice to prevent embryonic lethality associated with premature neural differentiation and angiogenic defects in the brain. Lethality is probably due to intraventricular haemorrhage which contributes to death in humans³⁰ and rodents²⁵. Angiogenesis associated with tumour growth and metastasis in adult animals is much more sensitive to Id dosage, as even partial loss of Id function results in profound defects in the neovascularization of tumours. Thus, we have shown for the first time to our knowledge that Id genes are required to maintain the timing of differentiation in mammalian development, and we have identified a role for Id proteins in angiogenesis which may be of clinical importance.

The premature neuronal differentiation in Id1–Id3 double knockout mice indicates that Id1 or Id3 is required to block the precisely timed expression and activation of positively acting bHLH proteins during murine development. There is a similar requirement for the negative bHLH regulator HES-1 (ref. 31). Redundancy in the Id family may account for the need to disrupt both genes and the timing of the observed effect: Id1, Id2 and Id3 are all expressed in dividing neuroblasts up to the time of premature neural differentiation (E11.5–E12.5), after which Id2 expression becomes localized to more mature neurons. The phenotype in the Id knockout mice is consistent with overexpression experiments showing that Id can inhibit both the activity and expression of tissue-restricted bHLH proteins by the sequestration of E protein heterodimerizing partners^{32,33}. Both types of regulation can be accomplished by the same biochemical mechanism given the transcription hierarchies established within the neural bHLH protein family: the expression of differentiation-inducing bHLH proteins is controlled by the activation of the upstream determination bHLH proteins. Indeed, the broader region of premature expression of NeuroD1 and MATH2 relative to the determination gene MATH3 in the ventral telencephalon is consistent with this hypothesis.

The hypocellularity or cavitation lesion in the ganglionic emi-

nences of Id1^{-/-}Id3^{-/-} mice is reminiscent of that associated with germinal matrix haemorrhage (GMH) affecting premature babies in which cavitation follows a haemorrhage³⁴. In Id1^{-/-}Id3^{-/-} embryos, however, the cavities appear before GMH, but at a time when red blood cells have escaped the vascular bed and have subsequently been engulfed by macrophages. Thus, cystic cavitation is probably related to vascular leakage and extravasation of fluid as in the class II αv -integrin-null embryos²⁵. We cannot exclude the possibility, however, that an alteration in cell migration is responsible for the cavitation. The dilation of the blood vessels within the ganglionic eminence may be due to a deficit in smooth muscle cells in vascular channels in the region of the malformation. Similar phenotypes have been reported in αv -integrin-null²⁵ and PDGFR β -null³⁵ embryos. The etiology of vascular rupture remains unclear but intracerebral haemorrhages also occur in αv -integrin-null embryos²⁵.

An aspect of the endothelial defect is probably cell autonomous, as reduced expression of VEGF and Flk-1 is observed in the mutant vascular malformation. This is consistent with the expression of Id1 and Id3 in the endothelial cells of wild-type animals. bHLH proteins probably contribute to the control of growth and maturation of

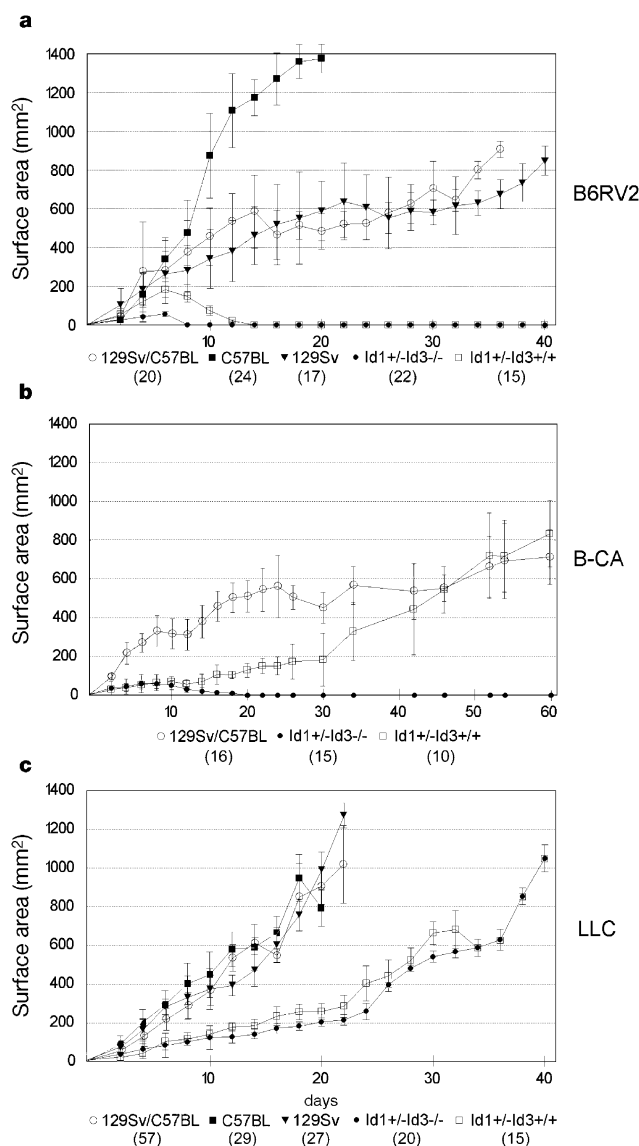


Figure 4 Tumour growth is inhibited in Id knockout mice. **a–c**, Wild-type and Id knockout mice were injected intradermally with tumour cell lines as indicated. Tumour surface area and the mean and standard deviations are shown. Number of animals is given in parentheses.

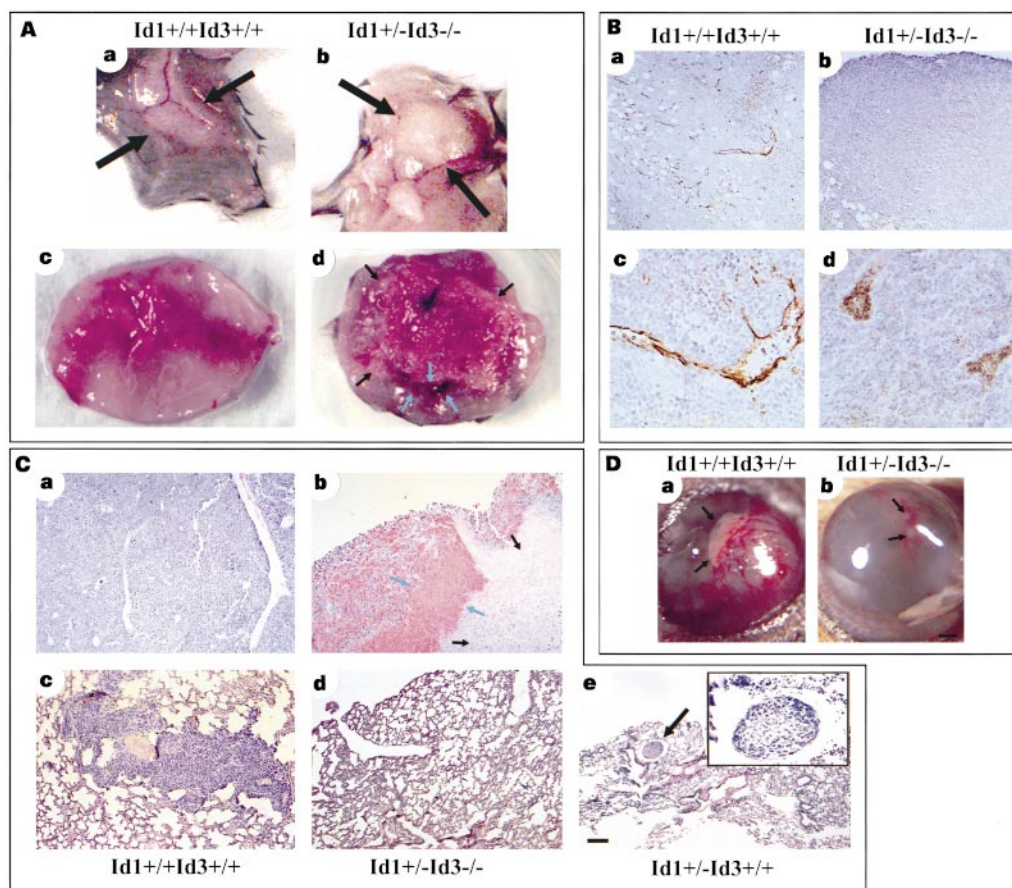


Figure 5 Decreased vascularization of xenografts and reduced metastases in $Id1^{+/-}Id3^{-/-}$ mutant mice. B6RV2 tumours 6 days after inoculation in wild-type and mutant mice shown by arrows in **Aa, b**; black arrows indicate necrosis and blue arrows outline haemorrhage in LLC tumours in mutant mice 40 days after injection (**Ad**) compared with wild-type (**Ac**). PECAM/CD31 stained blood vessels from day-8 wild-type (**Ba, c**) and mutant (**Bb, d**) B6RV2 tumours. Black arrows point to necrosis and blue arrows indicate

haemorrhage in the H&E stained mutant LLC tumours on day 20 after injection (**Cb**), as compared with the wild type (**Ca**). Metastases are seen in day 20 wild-type (**Cc**) but not in day-35 mutant lungs (**Cd**). A day-35 solitary lung nodule in the $Id1^{+/-}Id3^{+/+}$ mice is shown (**Ce**). **Da, b**, LLC (arrows) on day-8 post-intracorneal inoculation in wild-type (**Da**) and mutant mice (**Db**). Scale bar, 670 μ m (**A, Da, b**); 100 μ m (**Ba, b, C**); 25 μ m (**Bc, d, Ce** inset).

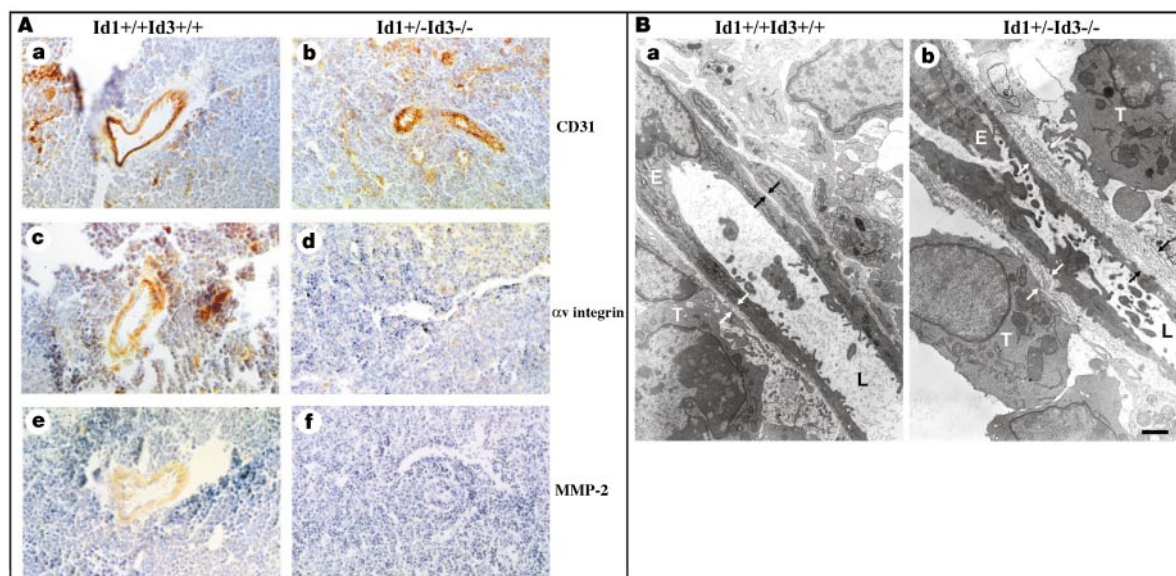


Figure 6 B6RV2 xenografts in $Id1^{+/-}Id3^{-/-}$ mice show reduced αv -integrin and MMP2 staining and a thickening of the extracellular matrix. CD31/PECAM (**Aa, b**), αv -integrin (**Ac, d**) and MMP2 (**Ae, f**) in adjacent wild-type and mutant tumour on day 8 after

inoculation. Arrows in electronographs on day 6 after injection outline the extracellular matrix of blood vessel in wild-type (**Ba**) and mutant tumours (**Bb**). L, lumen; E, endothelial cell; T, tumour cell. Scale bar, 25 μ m (**A**); 1.5 μ m (**B**).

endothelial cells; bHLH-EC2 (ref. 36) is expressed in endothelial cells and is a possible candidate for Id regulation. The localization of the endothelial defect specifically in the brain of Id1^{-/-}Id3^{-/-} embryos may be explained by functional redundancy: all three Id genes are expressed in the endothelium outside the brain, but only Id1 and Id3 are expressed in the endothelium within the brain. These observations highlight the notion that the full sphere of influence of the Id genes may not be determined until Id1–Id2–Id3 triple knockout animals are analysed.

Our data indicate that normal Id expression is also required to support angiogenesis in a primary tumour mass. Blood vessels in Id knockout animals lack the ability to branch and sprout and form normal calibre lumens during tumour progression. These processes depend on the proteolysis and remodelling of the extracellular matrix and, indeed, a pronounced thickening of the extracellular matrix surrounding endothelial cells in Id knockout animals occurred. Moreover, little if any α v β 3-integrin or its associated metalloproteinase MMP2 was detected in the stunted blood vessels in the residual tumour mass of Id1^{+/-}Id3^{-/-} mice. Thus, Id expression may directly or indirectly control the expression of α v β 3-integrin during tumour-induced angiogenesis and the recruitment and/or expression of MMP2 metalloproteinase which are required for tumour growth and metastasis^{37,38}. Notably, overexpression of Id activates the expression of a 120,000 dalton gelatinase in breast epithelia which enhances the invasiveness of these cells³⁹. Understanding the molecular mechanism of these effects is of obvious interest.

The Id proteins may be important new targets for anti-angiogenic drug design. As stated above, simply reducing the dosage of Id proteins seems to have a profound effect on the ability of tumours to grow and/or metastasize; therefore, a complete loss of Id function need not be obtained. In addition, the level of Id expression in almost all adult human tissues including resting endothelial cells (R.E. and D.L., unpublished data) is extremely low. Small-molecule inhibitors of Id might therefore be useful as anti-angiogenic drugs in the treatment of human cancers.

Although the analysis of tumour xenografts has identified important regulators of angiogenesis^{40,41}, tumours that develop in genetically manipulated mice present a more physiologically relevant model system. Whether loss of Id function affects the growth of such tumours is currently under investigation. □

Methods

Genotyping by PCR

Genomic DNA was obtained from mouse tail tips and yolk sacs as described⁴². PCR analyses were performed with primers specific for the wild-type and targeted alleles. Primer sequences for Id1 were pr-22 (common oligonucleotide; 5'-CCTCAGCGACACAA GATGCGATCG-3'), pr-k4 (wild-type oligonucleotide; 5'-GGTGTCTTTGAACGTTCT GAACC-3') and pr-pgk (mutant oligonucleotide; 5'-GCACGAGACTAGTGAGACGTG-3'). Primer sequences for Id3 were yz151 (common oligonucleotide; 5'-GTTTGAACA TAGTCTGCC-3'), yz170 (wild-type oligonucleotide; 5'-CACCGGCTCAGCGCCTTC AT-3'), and yz29 (mutant oligonucleotide; 5'-TCGACGCGATCGCCTTCTA-3'). PCR cycling conditions were 90 °C for 30 s, 57 °C for 30 s and 65 °C for 3 min, for 40 cycles. The amplified PCR products were analysed on 1% agarose gels to separate the wild-type (1.0 kb for Id1 and 2.0 kb for Id3) and targeted allele (0.8 kb for Id1 and 2.5 kb for Id3) fragments. PCR genotyping of α v-integrin-null mice was performed as described²⁵.

Morphological and histological analysis

Embryos were obtained from timed pregnancies, with noon of the plug date defined as E0.5. Embryos were fixed in 4% paraformaldehyde. Paraffin embedding was performed by dehydrating embryos through ethanol and HistoClear (National Diagnostics) before immersion in paraplast (Fisher Scientific). Sections of 6 or 7 μ m were stained with hematoxylin and eosin (H&E).

Tumours, lung and enucleated eyes were fixed, processed and stained with H&E as described above. The entire lung per animal was sectioned and analysed for metastases by two independent scorers. In Fig. 5b, blood vessels were counted in eight random 200 $\times$ fields and results from two independent scorers were averaged.

In situ hybridization

Embryos and tumours were fixed in 4% paraformaldehyde and embedded in paraffin. Sections (6 or 7 μ m) were processed for *in situ* hybridization with (α -³³P)UTP-labelled

antisense RNA probes, as described⁴³. Probe templates were provided by R. Kageyama (MATH-1, MATH-2 and MATH-3)⁴⁴, S. Tapscott (NeuroD2, NeuroD3)⁴⁵, J. Lee (NeuroD)⁴⁶, E. Lai (BF-1)²⁰, E. Boncinelli (Emx2)⁴⁷, P. Gruss (Pax6)²³, J. Rubenstein (Dlx2)²¹, and A. McMahon (Shh)⁴⁸. All Id templates were generated in our laboratory^{17,49}.

Whole-mount immunohistochemistry

Embryos were processed as described⁵⁰. Embryos were incubated with primary antibody (MEC 13.3 rat monoclonal anti-mouse PECAM-1 antibody; Pharmingen). Samples were incubated with biotinylated anti-rat antibody (Vector), and then with peroxidase-conjugated avidin (Vector). For colour detection, NiCl₂ and DAB were added to embryos.

Immunohistochemistry

For Ki67 immunohistochemistry, tissue antigens were unmasked and sections were incubated with monoclonal anti-mouse Ki67 antibody (NCL-Ki67-MM1; Novocastra Laboratories), followed by biotinylated anti-mouse antibody. The Histomouse-Sp Kit (Zymed Laboratories) was used. For all other antigens the following antibodies were used: MAP2 and p16, anti-rat MAP2 (clone HM-2; Sigma) or monoclonal anti-p16 (Santa Cruz) antibody; PECAM (CD31), MEC 13.3 monoclonal anti-mouse PECAM-1 antibody (Pharmingen); laminin, polyclonal anti-laminin antibody (Sigma); VEGF, anti-VEGF (Santa Cruz); Flk-1, anti-Flk-1 (Sigma); smooth muscle α -actin-anti- α -actin (Sigma); α v-integrin, polyclonal anti-mouse α v, integrin antibody (Chemicon); MMP2 (gelatinase-A), polyclonal anti-mouse MMP-2 antibody (Sigma). All sections were then incubated with the appropriate biotinylated secondary antibodies (Vector).

Cell lines

B6RV2, a murine leukemia/lymphoma cell line generated at Memorial Sloan-Kettering Cancer Center and LLC, obtained from American Type Culture Collection, were used. B-CA, a breast carcinoma cell line, was established from tumours generated by crossing Id1-null mice and mammary tumour virus–polyoma virus transgenic mice. All lines were maintained in DMEM with 10% fetal calf serum.

Tumour implantation in a murine model

Roughly 2×10^7 cells of each tumour cell line were injected intradermally in the right lower abdomen. Surface area was measured by two independent scorers (Dial Caliper, Science ware). For intravenous injection of tumour, 2×10^6 LCC cells were injected in the tail veins of anesthetized mice (2.5% Avertin). For the eye implantation studies, animals were anesthetized with 2.5% Avertin and propanecarboxylic acid hydrochloride ophthalmic solution (0.5%). Roughly 2×10^6 B6RV2, LLC cells or media alone were injected with a Hamilton syringe and needle with the assistance of a dissecting microscope (Zeiss Olympus) into the corneal layers of the eye.

Electron microscopy

Tissues were processed with 'yellowfix' (2.5% glutaraldehyde, 4% paraformaldehyde, 0.02% picric acid in 0.1 M Na-cacodylate). The samples were post-fixed with 1% osmium tetroxide–1.5% ferricyanide, dehydrated in ethanol and infiltrated with Spurr's Resin. Tissue blocks were trimmed with a diatome diamond knife (Diatome USA) on RMA MT700 ultramicrotome. Sections were contrasted with lead citrate and viewed on a JEOL 100CX-II electron microscope.

Received 24 April; accepted 6 September 1999.

- Norton, J. D. *et al.* Id helix-loop-helix proteins in cell growth and differentiation. *Trends Cell Biol.* **8**, 58–65 (1998).
- Lassar, A. B., Skapek, S. X. & Novitsch, B. Regulatory mechanisms that coordinate skeletal muscle differentiation and cell cycle withdrawal. *Curr. Opin. Cell Biol.* **6**, 788–794 (1994).
- Ron, D. & Habener, J. F. CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as dominant-negative inhibitor of gene transcription. *Genes Dev.* **6**, 439–453 (1992).
- Treacy, M. N., He, X. & Rosenfeld, M. G. I-POU: a POU-domain protein that inhibits neuron-specific gene activation. *Nature* **350**, 577–584 (1991).
- Yan, W. *et al.* High incidence of T-cell tumors in E2A-null mice and E2A/Id1 double-knockout mice. *Mol. Cell Biol.* **17**, 7317–7327 (1997).
- Deed, R. W., Jasiok, M. & Norton, J. D. Lymphoid-specific expression of the Id3 gene in hematopoietic cells. Selective antagonism of E2A basic helix-loop-helix protein associated with Id3-induced differentiation of erythroleukemia cells. *J. Biol. Chem.* **273**, 8278–8286 (1998).
- Ogata, T. *et al.* Bone morphogenetic protein 2 transiently enhances expression of a gene, Id (inhibitor of differentiation), encoding a helix-loop-helix molecule in osteoblast-like cells. *Proc. Natl Acad. Sci. USA* **90**, 9219–9222 (1993).
- Neuman, T. *et al.* Neuronal expression of regulatory helix-loop-helix factor Id2 gene in mouse. *Dev. Biol.* **160**, 186–195 (1993).
- Voronova, A. F. & Lee, F. The E2A and tal-1 helix-loop-helix proteins associated *in vivo* and are modulated by Id proteins during interleukin 6-induced myeloid differentiation. *Proc. Natl Acad. Sci. USA* **91**, 5952–5956 (1994).
- Heemskerk, M. H. *et al.* Inhibition of T cells and promotion of natural killer cell development by the dominant negative helix loop helix factor Id3. *J. Exp. Med.* **186**, 1597–1602 (1997).
- Martinsen, B. J. & Bronner-Fraser, M. Neural crest specification regulated by the helix-loop-helix repressor Id2. *Science* **281**, 988–991 (1998).
- Katagiri, T. *et al.* Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage [published erratum appears in *J. Cell Biol.* **128**, 713 (1995)]. *J. Cell Biol.* **127**, 1755–1766 (1994).

13. Yokota, Y. *et al.* Development of peripheral lymphoid organs and natural killer cells depends on the helix-loop-helix inhibitor Id2. *Nature* **397**, 702–706 (1999).
14. Pan, L. *et al.* Impaired immune responses and B cell proliferation in mice lacking the Id3 gene. *Mol. Cell Biol.* **19**, 5969–5980 (1999).
15. Duncan, M. *et al.* The gene for the helix-loop-helix protein, Id, is specifically expressed in neural precursors. *Dev. Biol.* **154**, 1–10 (1992).
16. Ellmeier, W. & Weith, A. Expression of the helix-loop-helix gene Id3 during murine embryonic development. *Dev. Dyn.* **203**, 163–173 (1995).
17. Jen, Y., Manova, K. & Benezra, R. Each member of the Id gene family exhibits a unique expression pattern in mouse gastrulation and neurogenesis. *Dev. Dyn.* **208**, 92–106 (1997).
18. Risau, W. Mechanisms of angiogenesis. *Nature* **386**, 671–674 (1997).
19. Lee, J. E. Basic helix-loop-helix genes in neural development. *Curr. Opin. Neurobiol.* **7**, 13–20 (1997).
20. Xuan, S. *et al.* Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres. *Neuron* **14**, 1141–1152 (1995).
21. Price, M. *et al.* A mouse gene related to Distal-less shows a restricted expression in the developing forebrain. *Nature* **351**, 748–751 (1991).
22. Simeone, A. *et al.* A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid class and demarcates anterior neuroectoderm in the gastrulating mouse embryo. *EMBO J.* **12**, 2735–2747 (1993).
23. Walther, C. & Gruss, P. Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development* **113**, 1435–1449 (1991).
24. Shimamura, K. *et al.* Longitudinal organization of the anterior neural plate and neural tube. *Development* **121**, 3923–3933 (1995).
25. Bader, B. L. *et al.* Extensive vasculogenesis, angiogenesis, and organogenesis precede lethality in mice lacking all alpha v integrins. *Cell* **95**, 507–519 (1998).
26. Bain, G. *et al.* E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements [see comments]. *Cell* **79**, 885–892 (1994).
27. Zhuang, Y., Soriano, P. & Weintraub, H. The helix-loop-helix gene E2A is required for B cell formation. *Cell* **79**, 875–884 (1994).
28. Erickson, J. W. & Cline, T. W. Molecular nature of the *Drosophila* sex determination signal and its link to neurogenesis. *Science* **251**, 1071–1074 (1991).
29. Brooks, P. C. *et al.* Localization of matrix metalloproteinase MMP-2 to the surface of invasive cells by interaction with integrin alpha v beta 3. *Cell* **85**, 683–693 (1996).
30. Perlman, J. M. & Volpe, J. J. Intraventricular hemorrhage in extremely small premature infants. *Am. J. Dis. Child.* **140**, 1122–1124 (1986).
31. Ishibashi, M. *et al.* Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev.* **9**, 3136–3148 (1995).
32. Jen, Y., Weintraub, H. & Benezra, R. Overexpression of Id protein inhibits the muscle differentiation program: *in vivo* association of Id with E2A protein. *Genes Dev.* **6**, 1466–1479 (1992).
33. Melnikova, I. N. & Christy, B. A. Muscle cell differentiation is inhibited by the helix-loop-helix protein Id3. *Cell Growth Differ.* **7**, 1067–1079 (1996).
34. Sherwood, A., Hopp, A. & Smith, J. F. Cellular reactions to subependymal plate haemorrhage in the human neonate. *Neuropathol. Appl. Neurobiol.* **4**, 245–261 (1978).
35. Lindahl, P. *et al.* Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277**, 242–245 (1997).
36. Quertermous, E. E. *et al.* Cloning and characterization of a basic helix-loop-helix protein expressed in early mesoderm and the developing somites. *Proc. Natl Acad. Sci. USA* **91**, 7066–7070 (1994).
37. Brooks, P. C. *et al.* Integrin alpha v beta 3 antagonists promote tumor regression by inducing apoptosis of angiogenic blood vessels. *Cell* **79**, 1157–1164 (1994).
38. Zetter, B. R. Angiogenesis and tumor metastasis. *Annu. Rev. Med.* **49**, 407–424 (1998).
39. Desprez, P. Y. *et al.* A novel pathway for mammary epithelial cell invasion induced by the helix-loop-helix protein Id-1. *Mol. Cell Biol.* **18**, 4577–4588 (1998).
40. O'Reilly, M. S. *et al.* Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. *Cell* **88**, 277–285 (1997).
41. Boehm, T. *et al.* Antiangiogenic therapy of experimental cancer does not induce acquired drug resistance [see comments]. *Nature* **390**, 404–407 (1997).
42. Hogan, B. L. *et al.* Small eyes (Sey): a homozygous lethal mutation on chromosome 2 which affects the differentiation of both lens and nasal placodes in the mouse. *J. Embryol. Exp. Morphol.* **97**, 95–110 (1986).
43. Manova, K. *et al.* Apoptosis in mouse embryos: elevated levels in pregastrulae and in the distal anterior region of gastrulae of normal and mutant mice. *Dev. Dyn.* **213**, 293–308 (1998).
44. Akazawa, C. *et al.* A mammalian helix-loop-helix factor structurally related to the product of *Drosophila* proneural gene atonal is a positive transcriptional regulator expressed in the developing nervous system. *J. Biol. Chem.* **270**, 8730–8738 (1995).
45. McCormick, M. B. NeuroD2 and neuroD3: distinct expression patterns and transcriptional activation potentials within the neuroD gene family. *Mol. Cell Biol.* **16**, 5792–5800 (1996).
46. Lee, J. E. *et al.* Conversion of *Xenopus* ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science* **268**, 836–844 (1995).
47. Simonson, M. S., Rooney, A. & Herman, W. H. Expression and differential regulation of Id1, a dominant negative regulator of basic helix-loop-helix transcription factors, in glomerular mesangial cells. *Nucleic Acids Res.* **21**, 5767–5774 (1993).
48. Echelard, Y. *et al.* Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417–1430 (1993).
49. Jen, Y., Manova, K. & Benezra, R. Expression patterns of Id1, Id2, and Id3 are highly related but distinct from that of Id4 during mouse embryogenesis. *Dev. Dyn.* **207**, 235–252 (1996).
50. Winnier, G. E., Hargett, L. & Hogan, B. L. The winged helix transcription factor MFH1 is required for proliferation and patterning of paraxial mesoderm in the mouse embryo. *Genes Dev.* **11**, 926–940 (1997).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank F. Giancotti, E. Lai and K. Hajjar for critical reading of the manuscript and members of the Benezra lab for helpful comments throughout the course of the work. S. Kerns, K. Witty-Blease, J. Waka and A. Nicola of the Molecular Cytology Core Facility at SKI provided invaluable technical assistance. We thank R. Alani for establishing the B-CA cell line, W. Mark for the B6RV2 cell line, L. Cohen-Gould for help with electron microscopy, H. Nguyen and H. Gultekin for histological consultation and S. Doshi for assistance with eye surgery. R.B. thanks J. Benezra for unwavering support and all participants of the Annual Harold Weintraub Memorial Symposium in Aspen Colorado for their insightful comments. This work was supported by grants from the Children's Brain Tumor Foundation and from the NIH (K12) (D.L.), a grant from the NIH Medical Scientist Training Program (A.Y.), a grant from NIH Cancer Center Support (K.M.) and grants from the NIH and NSF (R.B.).

Correspondence and requests for materials should be addressed to R.B. (e-mail: r-benezra@ski.mskcc.org).

Bright rings around sunspots

M. P. Rast*, P. A. Fox*, H. Lin†, B. W. Lites*, R. W. Meisner* & O. R. White*

* High Altitude Observatory, National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307, USA

† National Solar Observatory, Sunspot, New Mexico 88349, USA

There are two possible explanations for why sunspots are dark: the partial suppression by the sunspot magnetic fields of convective energy transport from the underlying layers¹, or the removal of energy from the sunspot by enhanced hydromagnetic wave radiation². Both processes would reduce the energy emitted radiatively. The first explanation is currently favoured³, and predicts that the blocked energy should show up as a bright ring around the spot², with the actual brightness of the ring sensitive to details of solar convective transport and sunspot structure⁴. Previous searches⁵ for these bright rings were inconclusive because of the presence of bright, vertical magnetic flux tubes near the spots, and a lack of sufficient precision in the observations. Here we report high-photometric-precision obser-

vations of bright rings around eight sunspots. The rings are about 10 K warmer than the surrounding photosphere and extend at least one sunspot radius out from the penumbra. About 10% of the radiative energy missing from the sunspots is emitted through the bright rings. We also report observations of a second set of sunspots, for which simultaneous magnetic field measurements demonstrate that the rings are not associated with vertical flux tubes.

Twelve sunspots were selected, on the basis of their symmetric appearance, from full disk images taken by the Precision Solar Photometric Telescope (PSPT) at Mauna Loa Solar Observatory. A set of images of each spot was chosen using the following uniformly applied criteria: nearly simultaneous (in the same hourly observing sequence) red-band and blue-band images of the spot exist, the sunspot is near the disk centre ($\mu \geq 0.8$, where μ is the cosine of the heliocentric viewing angle), and the transparency and seeing are good (disk centre intensity and r.m.s. fluctuations in the quiet Sun exceed or equal threshold values; the r.m.s. intensity fluctuation thresholds employed were 1% in the blue and 0.5% in the red). Of the 12 sunspots originally selected, 10 had at least one set of images meeting these criteria (8 had more than one set of such images).

The best example from the 12-sunspot sample shows continuum intensity enhancements surrounding the spot of the order of 1% (relative to the mean quiet Sun), with larger enhancements in the blue images than in the red (Fig. 1). This is consistent with an elevation in temperature of the order of 10 K, with very good agreement between the azimuthally averaged ring-brightness-temperature perturbations measured (relative to the disk centre values⁶) in the two colour bands separately (Fig. 2a). The temperature enhancement in the ring is much larger than either the

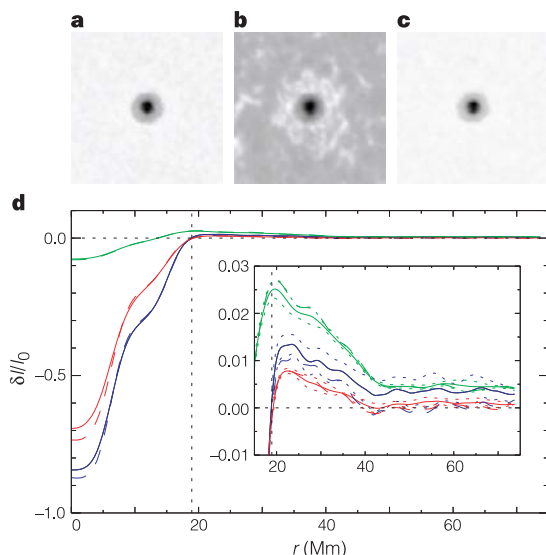


Figure 1 Bright rings are seen in all wavelength bands measured. **a–c**, Intensity fluctuations $\delta I/I_0$ through the blue, Ca II K, and red filters of the Precision Solar Photometric Telescope (PSPT) in a 150 Mm² area surrounding a large isolated sunspot. **d**, Azimuthal averages of these intensity fluctuations as a function of radial distance r from the spot centre: blue filter (blue line), Ca II K (green line), red filter (red line); Ca II K fluctuations are plotted reduced by a factor of 10. Inset, detail of the region immediately outside the penumbral boundary. Dashed curves plot the azimuthal averages obtained from the single images shown, solid curves plot averages of the azimuthal curves over the image set defined in the text, and dotted curves indicate the r.m.s. deviations from these averages. We note the rather odd network appearance, reduced in size from typical supergranular network scales, of the moderate intensity Ca II K emission surrounding the spot, suggesting enhanced convection in the ring region. The PSPT 2,048 × 2,048 pixel full-disk images were taken at wavelengths of 409 nm (continuum blue), 607 nm (continuum red) and 393 nm (line centre Ca II K). Variations in the detector sensitivity ('flat-fields') were computed using a modified version of the algorithm of Kuhn, Lin and Loran¹⁰, applied to 16 offset images of the Sun. The raw images were divided by the 'flat-field' and a two-dimensional spatial linear gradient plus limb-darkening (intensity variation due to solar sphericity) function I_0 was then fitted to the full-disk quiet-Sun pixels. Residual intensity fluctuations $\delta I/I_0 = (I - I_0)/I_0$ were computed at every point, and subimages containing the sunspots were projected to a normal viewing angle.

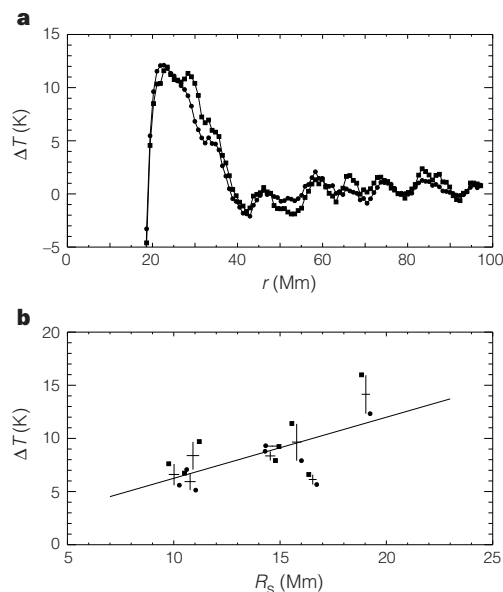


Figure 2 Ring temperature enhancements are of the order of 10 K and scale with sunspot size. **a**, Azimuthally averaged sunspot ring-brightness-temperature perturbations ΔT plotted as a function of distance from the spot centre r for the red (circles) and blue (squares) images shown in Fig. 1a and c. **b**, Peak brightness-temperature perturbations plotted as a function of sunspot radius R_s for eight different sunspots. The sunspot radius is defined to be the location at which the azimuthally averaged $\delta I/I_0$ crosses zero (the quiet-Sun value) at the edge of the penumbra. The centre of the crosses in **b** indicate the average between the red and blue brightness temperatures and spot radii. A linear fit to these average values is also plotted (solid line). There is an apparent correlation between sunspot radius and ring brightness or temperature, with larger sunspots having brighter rings.

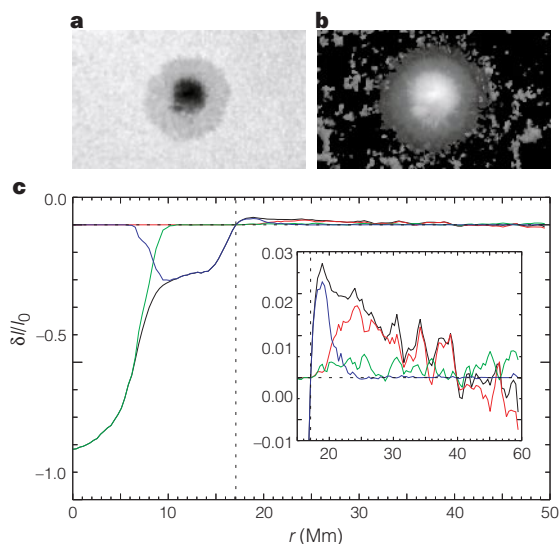


Figure 3 The enhanced brightness in the ring originates mainly in regions of very weak magnetic field. Continuum intensity (**a**) and magnetic field strength (**b**) in and around a sunspot are shown as determined by the Advanced Stokes Polarimeter (ASP), showing high-intensity and strong-field regions as bright areas. For these observations the ASP measured the continuum intensity near 630 nm and full Stokes polarization profiles of the Fe I 630.15 nm and 630.25 nm lines. A spectral resolution of ~ 30 mÅ was used, slit width of 0.6 arcsec, mapping the sunspot and its surroundings with 0.75 arcsec steps of the slit. Photospheric vector magnetic fields were inferred¹¹ at all spatial locations with an integrated polarization $\geq 0.4\%$, providing magnetic field strength and direction along with continuum intensity at each image location. The darkest pixels in the field strength map correspond to those for which no inversion was attempted. The mean "apparent flux density"¹² for those pixels is ≈ 10 –20 G. **c**, Azimuthal averages of the continuum intensity fluctuations $\delta I/I_0$ as a function of distance r from the sunspot centre and colour coded by contribution from pixels of measured magnetic field zenith angle. The black curve plots the total azimuthally averaged intensity from all pixels, and the coloured curves plot the contributions to this total from pixels with measured vertical magnetic field (green line), measured horizontal field (blue line), or weak or zero field (red line). Inset, detail of the region immediately outside the penumbral boundary. The vertical magnetic field regions contribute negligibly to the sunspot bright ring, with most of the ring intensity excess originating in pixels of very weak field.

difference between the blue and red measurements, or the fluctuations in either measurement individually outside the ring region. The ring extends about two sunspot radii outward from the spot centre (with an area about three times that of the spot); this implies that the emitted energy excess is about 10% of the energy deficit of the spot (the spot being on average about 40% dimmer than its surroundings). Peak ring-brightness-temperatures were similarly evaluated for the remaining spots for which more than one image in each colour band was available. Not only do sunspot bright rings exist, but there is an apparent correlation between sunspot radius and ring temperature (Fig. 2b). Larger spots have brighter, hotter rings surrounding them. As ring size also increases with sunspot size (always extending about two sunspot radii outward from the spot centre), this correlation implies either that larger sunspots or their penumbra are shallower (assuming conventional turbulent diffusion models are correct), or that ordered convective flows of gas surrounding larger spots are more vigorous, transporting more of the blocked flux to the surface.

It is important to determine whether the continuum intensity excess measured is diffuse thermal emission or due to hot, bright, vertical flux tubes surrounding the sunspot. We address this issue directly using precision spectro-polarimetric observations of sunspots taken with the Advanced Stokes Polarimeter (ASP) (Fig. 3a, b). Two of four sunspots examined showed strong evidence of a bright ring in continuum intensity. Contributions to the ring intensity as a function of field zenith angle (Fig. 3c) clearly demonstrate that sites of vertical magnetic field (with field zenith angles $>135^\circ$ or $<45^\circ$ and mean field strengths of about 1,000 G) make a negligible contribution to the total ring intensity. Horizontal field locations (with zenith angles $>45^\circ$ and $<135^\circ$ and mean field strengths of about 350 G) immediately outside the continuum intensity penumbral boundary contribute most of the intensity excess of the inner ring. It is not known however whether this horizontal field is causally related to the intensity enhancement. It could be an indication of siphon flow around the sunspot, but is more likely to be overlying canopy field unrelated to the continuum enhancement. Nearly all of the remaining ring intensity excess originates in pixels of very weak polarization (mean field strengths ≈ 10 –20 G). Although possible contributions from vertical magnetic elements below the resolution limit of the observations cannot be ruled out entirely, it should be noted that as the polarization threshold for field determination is lowered, the amount of vertical field in the sunspot ring is not

significantly altered. Instead, additional weak horizontal canopy fields are measured.

Thus, isolated sunspots are seen to be commonly surrounded by a ring of enhanced radiation, the origin of which is probably not bright vertical magnetic elements (plage field), but the re-emergence of heat blocked by magnetic inhibition of convective transport in the spot itself. The magnitude of the enhancement suggests two possibilities. Either sunspots are a shallow phenomenon (a sunspot 40 Mm in diameter being less than ~ 3 Mm deep⁷ or having a penumbra ~ 1 Mm deep⁴) even though their ultimate origin is thought to be in the solar dynamo acting near the base of the convection zone, or magnetohydrodynamic flows surrounding sunspots transport heat to the surface more efficiently than is expected on the basis of turbulent diffusion models⁸. Such flows may play an important role in sunspot formation and maintenance, and could be related to observed sunspot moats⁹. □

Received 21 June; accepted 20 August 1999.

1. Biermann, L. Der gegenwärtige Stand der Theorie konvektiver Sonnenmodelle. *Vierteljahrsh. Astron. Gesellsch.* **76**, 194–200 (1941).
2. Parker, E. N. The nature of the sunspot phenomenon. I. Solutions of the heat transport equation. *Sol. Phys.* **36**, 249–274 (1974).
3. Thomas, J. H. & Weiss, N. O. in *Sunspots: Theory and Observations* (eds Thomas, J. H. & Weiss, N. O.) 3–59 (Kluwer, Dordrecht, 1992).
4. Spruit, H. C. Heat flow near obstacles in the solar convection zone. *Sol. Phys.* **55**, 3–34 (1977).
5. Fowler, L. A., Foukal, P. & Duvall, T. Jr Sunspot bright rings and the thermal diffusivity of solar convection. *Sol. Phys.* **84**, 33–44 (1983).
6. Neckel, H. & Labs, D. The solar radiation between 3300 and 12500 Å. *Sol. Phys.* **90**, 205–258 (1984).
7. Spruit, H. C. The flow of heat near a starspot. *Astron. Astrophys.* **108**, 356–360 (1982).
8. Fox, P. A., Sofia, S. & Chan, K. L. Convective flows around sunspot-like objects. *Sol. Phys.* **135**, 15–42 (1991).
9. Brickhouse, N. S. & LaBonte, B. J. Mass and energy flow near sunspots I. Observations of moat properties. *Sol. Phys.* **115**, 43–60 (1988).
10. Kuhn, J. R., Lin, H. & Loran, D. Gain calibrating nonuniform image-array data using only the image data. *Publ. Astron. Soc. Pacif.* **103**, 1097–1108 (1991).
11. Skumanich, A. & Lites, B. W. Stokes profile analysis and vector magnetic fields. I. Inversion of photospheric lines. *Astrophys. J.* **322**, 473–482 (1987).
12. Lites, B. W., Rutten, R. J. & Berger, T. E. Dynamics of the solar chromosphere. II. Ca II H₂ and K₂ grains versus internetwork fields. *Astrophys. J.* **517**, 1013–1033 (1999).

Acknowledgements

We thank J. Kuhn, K. MacGregor, C. Rice, P. Seagraves, A. Skumanich, K. Streader and M. Woodard. The National Center for Atmospheric Research and the National Solar Observatory are sponsored by the National Science Foundation.

Correspondence and requests for materials should be addressed to M.P.R. (e-mail: mprast@ucar.edu).

Wave-particle duality of C₆₀ molecules

Markus Arndt, Olaf Nairz, Julian Vos-Andreae, Claudia Keller, Gerbrand van der Zouw & Anton Zeilinger

Institut für Experimentalphysik, Universität Wien, Boltzmannngasse 5, A-1090 Wien, Austria

Quantum superposition lies at the heart of quantum mechanics and gives rise to many of its paradoxes. Superposition of de Broglie matter waves¹ has been observed for massive particles such as electrons², atoms and dimers³, small van der Waals clusters⁴, and neutrons⁵. But matter wave interferometry with larger objects has remained experimentally challenging, despite the development of powerful atom interferometric techniques for experiments in fundamental quantum mechanics, metrology and lithography⁶. Here we report the observation of de Broglie wave interference of C₆₀ molecules by diffraction at a material absorption grating. This molecule is the most massive and complex object in which wave behaviour has been observed. Of particular interest is the fact that C₆₀ is almost a classical body, because of its many excited internal degrees of freedom and their possible couplings to the environment. Such couplings are essential for the appearance of decoherence^{7,8}, suggesting that interference experiments with large molecules should facilitate detailed studies of this process.

When considering de Broglie wave phenomena of larger and more complex objects than atoms, fullerenes come to mind as suitable candidates. After their discovery⁹ and the subsequent invention of efficient mass-production methods¹⁰, they became easily available. In our experiment (see Fig. 1) we use commercial, 99.5% pure, C₆₀ fullerenes (Dynamic Enterprises Ltd, Twyford, UK) which were sublimated in an oven at temperatures between 900 and 1,000 K. The emerging molecular beam passed through two collimation slits, each about 10 µm wide, separated by a distance of 1.04 m. Then it traversed a free-standing nanofabricated SiN_x grating¹¹ consisting of nominally 50-nm-wide slits with a 100-nm period.

At a further distance of 1.25 m behind the diffraction grating, the interference pattern was observed using a spatially resolving detector. It consisted of a beam from a visible argon-ion laser (24 W all lines), focused to a gaussian waist of 8 µm width (this is the size required for the light intensity to drop to 1/e² of that in the centre of the beam). The light beam was directed vertically, parallel both to the lines of the diffraction grating and to the collimation slits. By using a suitable mirror assembly, the focus could be scanned with micrometre resolution across the interference pattern. The absorbed light then ionized the C₆₀ fullerenes via heating and subsequent thermal emission of electrons¹². The detection region

was found to be smaller than 1 mm in height, consistent with a full Rayleigh length of 800 µm and the strong power dependence of this ionization process. A significant advantage of the thermionic mechanism is that it does not detect any of the residual gases present in the vacuum chamber. We could thus achieve dark count rates of less than one per second even under moderately high vacuum conditions (5 × 10⁻⁷ mbar). The fullerene ions were then focused by an optimized ion lens system, and accelerated to a BeCu conversion electrode at -9 kV where they induced the emission of electrons which were subsequently amplified by a Channeltron detector.

Alignment is a crucial part of this experiment. In order to be able to find the beam in the first place, our collimation apertures are movable piezo slits that can be opened from 0 to 60 µm (in the case of the first slit) and from 0 to 200 µm (for the second slit). The vacuum chamber is rigidly mounted on an optical table together with the ionizing laser, in order to minimize spatial drifts.

The effect of gravity also had to be considered in our set-up. For the most probable velocity (220 m s⁻¹), the fullerenes fall by 0.7 mm while traversing the apparatus. This imposes a constraint on the maximum tilt that the grating may have with respect to gravity. As a typical diffraction angle into the first-order maximum is 25 µrad, one can tolerate a tilt angle of (at most) about one mrad before molecules start falling from one diffraction order into the trajectory of a neighbouring order of a different velocity class. The experimental curves start to become asymmetric as soon as the grating tilt deviates by more than 500 µrad from its optimum vertical orientation.

The interference pattern of Fig. 2a clearly exhibits the central maximum and the first-order diffraction peaks. The minima between zeroth and first orders are well developed, and are due to destructive interference of C₆₀ de Broglie waves passing through neighbouring slits of the grating. For comparison, we show in Fig. 2b the profile of the undiffracted collimated beam. The velocity distribution has been measured independently by a time-of-flight method; it can be well fitted by $f(v) = v^3 \exp(-(v - v_0)^2/v_m^2)$, with $v_0 = 166 \text{ m s}^{-1}$ and $v_m = 92 \text{ m s}^{-1}$ as expected for a transition between a maxwellian effusive beam and a supersonic beam¹³. The most probable velocity was $v = 220 \text{ m s}^{-1}$, corresponding to a de Broglie wavelength of 2.5 pm. The full-width at half-maximum was as broad as 60%, resulting in a longitudinal coherence length of about 5 pm.

The essential features of the interference pattern can be understood using standard Kirchhoff diffraction theory¹⁴ for a grating with a period of 100 nm, by taking into account both the finite width of the collimation and the experimentally determined velocity distribution. The parameters in the fit were the width of the collimation, the gap width s_0 of a single slit opening, the effective beam width of the detection laser and an overall scaling factor. This model, assuming all grating slits to be perfect and identical, reproduces very well the central peak of the interference pattern shown in Fig. 2a, but does not fit the 'wings' of this pattern.

Agreement with the experimental data, including the 'wings' in

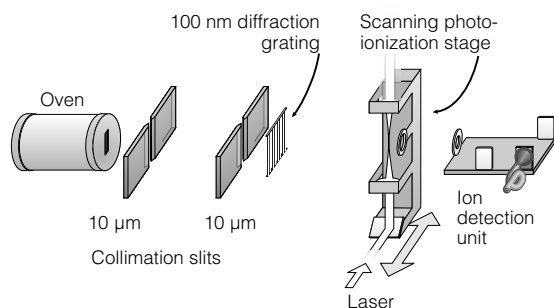


Figure 1 Diagram of the experimental set-up (not to scale). Hot, neutral C₆₀ molecules leave the oven through a nozzle of 0.33 mm × 1.3 mm × 0.25 mm (width × height × depth), pass through two collimating slits of 0.01 mm × 5 mm (width × height) separated by 1.04 m, traverse a SiN_x grating (period 100 nm) 0.1 m after the second slit, and are detected via thermal ionization by a laser 1.25 m behind the grating. The ions are then accelerated and directed towards a conversion electrode. The ejected electrons are subsequently counted by a Channeltron electron multiplier. The laser focus can be reproducibly scanned transversely to the beam with 1-µm resolution.

Fig. 2a, can be achieved by allowing for a gaussian variation of the slit widths over the grating, with a mean open gap width centred at $s_0 = 38$ nm with a full-width at half-maximum of 18 nm. That best-fit value for the most probable open gap width s_0 is significantly smaller than the 55 ± 5 nm specified by the manufacturer (T. A. Savas and H. Smith, personal communication). This trend is consistent with results obtained in the diffraction of noble gases and He clusters, where the apparently narrower slit was interpreted as being due to the influence of the van der Waals interaction with the SiN_x grating during the passage of the molecules¹⁵. This effect is expected to be even more pronounced for C_{60} molecules owing to their larger polarizability. The width of the distribution seems also justified in the light of previous experiments with similar gratings: both the manufacturing process and adsorbents could account for this fact (ref. 16, and T. A. Savas and H. Smith, personal communication). Recently, we also observed interference of C_{70} molecules.

Observation of quantum interference with fullerenes is interesting for various reasons. First, the agreement between our measured and calculated interference contrast suggests that not only the highly symmetric, isotopically pure $^{12}\text{C}_{60}$ molecules contribute to the interference pattern but also the less symmetric isotopomeric variants $^{12}\text{C}_{59}^{13}\text{C}$ and $^{12}\text{C}_{58}^{13}\text{C}_2$ which occur with a total natural abundance of about 50%. If only the isotopically pure $^{12}\text{C}_{60}$ molecules contributed to the interference, we would observe a much larger background.

Second, we emphasize that for calculating the de Broglie wavelength, $\lambda = h/Mv$, we have to use the complete mass M of the object. Thus, each C_{60} molecule acts as a whole undivided particle during its centre-of-mass propagation.

Last, the rather high temperature of the C_{60} molecules implies broad distributions, both of their kinetic energy and of their internal energies. Our good quantitative agreement between experiment and theory indicates that the latter do not influence the observed coherence. All these observations support the view that each C_{60} molecule interferes with itself only.

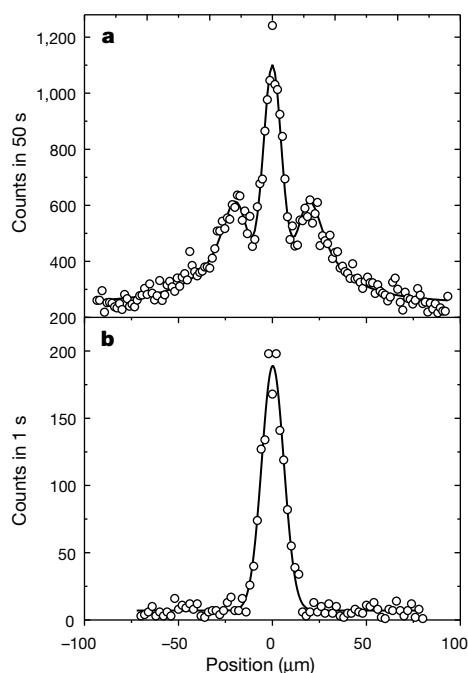


Figure 2 Interference pattern produced by C_{60} molecules. **a**, Experimental recording (open circles) and fit using Kirchhoff diffraction theory (continuous line). The expected zeroth and first-order maxima can be clearly seen. Details of the theory are discussed in the text. **b**, The molecular beam profile without the grating in the path of the molecules.

In quantum interference experiments, coherent superposition only arises if no information whatsoever can be obtained, even in principle, about which path the interfering particle took. Interaction with the environment could therefore lead to decoherence. We now analyse why decoherence has not occurred in our experiment and how modifications of our experiment could allow studies of decoherence using the rich internal structure of fullerenes.

In an experiment of the kind reported here, 'which-path' information could be given by the molecules in scattering or emission processes, resulting in entanglement with the environment and a loss of interference. Among all possible processes, the following are the most relevant: decay of vibrational excitations via emission of infrared radiation, emission or absorption of thermal blackbody radiation over a continuous spectrum, Rayleigh scattering, and collisions.

When considering these effects, one should keep in mind that only those scattering processes which allow us to determine the path of a C_{60} molecule will completely destroy in a single event the interference between paths through neighbouring slits. This requires $\lambda \ll d$; that is, the wavelength λ of the incident or emitted radiation has to be smaller than the distance d between neighbouring slits, which amounts to 100 nm in our experiment. When this condition is not fulfilled decoherence is however also possible via multi-photon scattering^{7,8,17}.

At $T \approx 900$ K, as in our experiment, each C_{60} molecule has on average a total vibrational energy of $E_v \approx 7$ eV (ref. 18) stored in 174 vibrational modes, four of which may emit infrared radiation at $\lambda_{\text{vib}} \approx 7\text{--}19$ μm (ref. 10) each with an Einstein coefficient of $A_k \approx 100$ s^{-1} (ref. 18). During its time of flight from the grating towards the detector ($\tau \approx 6$ ms) a C_{60} molecule may thus emit on average 2–3 such photons.

In addition, hot C_{60} has been observed¹⁹ to emit continuous blackbody radiation, in agreement with Planck's law, with a measured integrated emissivity of $\epsilon \approx 4.5 (\pm 2.0) \times 10^{-5}$ (ref. 18). For a typical value of $T \approx 900$ K, the average energy emitted during the time of flight can then be estimated as only $E_{\text{bb}} \approx 0.1$ eV. This corresponds to the emission of (for example) a single photon at $\lambda \approx 10$ μm . Absorption of blackbody radiation has an even smaller influence as the environment is at a lower temperature than the molecule. Finally, since the mean free path for neutral C_{60} exceeds 100 m in our experiment, collisions with background molecules can be neglected.

As shown above, the wavelengths involved are too large for single photon decoherence. Also, the scattering rates are far too small to induce sufficient phase diffusion. This explains the decoupling of internal and external degrees of freedom, and the persistence of interference in our present experiment.

A variety of unusual decoherence experiments would be possible in a future extension of the experiment, using a large-area interferometer. A three-grating Mach–Zehnder interferometer⁶ seems to be a particularly favourable choice, since for a grating separation of up to 1 m we will have a molecular beam separation of up to 30 μm , much larger than the wavelength of a typical thermal photon. In this case, the environment obtains 'which-path' information even through a single thermal photon, and the interference contrast should thus be completely destroyed. The parameters that could be controlled continuously in such an experiment would then be the internal temperature of the fullerenes, the temperature of the environment, the intensity and frequency of external laser radiation, the interferometer size, and the background pressure of various gases.

An improved interferometer could have other applications. For example, in contrast to previous atom-optical experiments^{20–22} which were limited to the interaction with only a few lines in the whole spectrum, interferometry with fullerenes would enable us to study these naturally occurring and ubiquitous thermal processes and wavelength-dependent decoherence mechanisms for (we

believe) the first time. Another possible application of molecule interferometry is precision metrology; the improved interferometer could be used to measure molecular polarizabilities⁶. Moreover, it might be possible to nanofabricate SiC patterns on Si substrates using C₆₀ interferometry.

Furthermore, we note the fundamental difference between isotopically pure C₆₀, which should exhibit bosonic statistics, and the isotopomers containing one ¹³C nucleus, which should exhibit fermionic statistics. We intend to explore the possibility of observing this feature, for example by showing the different rotational symmetry between the two species in an interferometer^{23,24}.

In our experiment, the de Broglie wavelength of the interfering fullerenes is already smaller than their diameter by a factor of almost 400. It would certainly be interesting to investigate the interference of objects the size of which is equal to or even bigger than the diffracting structure. Methods analogous to those used for the present work, probably extended to the use of optical diffraction structures, could also be applied to study quantum interference of even larger macromolecules or clusters, up to small viruses^{25,26}. □

Received 30 June; accepted 2 September 1999.

1. de Broglie, L. Waves and quanta. *Nature* **112**, 540 (1923).
2. Davisson, C. J. & Germer, L. H. The scattering of electrons by a single crystal of nickel. *Nature* **119**, 558–560 (1927).
3. Estermann, I. & Stern, O. Beugung von Molekularstrahlen. *Z. Phys.* **61**, 95–125 (1930).
4. Schöllkopf, W. & Toennies, J. P. Nondestructive mass selection of small van der Waals clusters. *Science* **266**, 1345–1348 (1994).
5. Halban, H. v. Jr & Preiswerk, P. Preuve expérimentale de la diffraction des neutrons. *C.R. Acad. Sci.* **203**, 73–75 (1936).
6. Berman, P. (ed.) *Atom Interferometry* (Academic, 1997).
7. Zurek, W. H. Decoherence and the transition from quantum to classical. *Phys. Today* 36–44 (October 1991).
8. Giulini, D. et al. *Decoherence and the Appearance of the Classical World in Quantum Theory* (Springer, Berlin, 1996).
9. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. C₆₀: buckminsterfullerene. *Nature* **318**, 162–166 (1985).
10. Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. A new form of carbon. *Nature* **347**, 354–358 (1990).
11. Savas, T. A., Shah, S. N., Schattenburg, M. L., Carter, J. M. & Smith, H. I. Achromatic interferometric lithography for 100-nm-period gratings and grids. *J. Vac. Sci. Technol. B* **13**, 2732–2735 (1995).
12. Ding, D., Huang, J., Compton, R. N., Klotz, C. E. & Haufler, R. E. cw laser ionization of C₆₀ and C₇₀. *Phys. Rev. Lett.* **73**, 1084–1087 (1994).
13. Scyles, G. (ed.) *Atomic and Molecular Beam Methods* Vol. 1 (Oxford Univ. Press, 1988).
14. Born, M. & Wolf, E. *Principles of Optics* (Pergamon, Oxford, 1984).
15. Grisenti, R. E., Schöllkopf, W., Toennies, J. P., Hegerfeldt, G. C. & Köhler, T. Determination of atom-surface van der Waals potentials from transmission-grating diffraction intensities. *Phys. Rev. Lett.* **83**, 1755–1758 (1999).
16. Grisenti, R. E. et al. He atom diffraction from nanostructure transmission gratings: the role of imperfections. *Phys. Rev. A* (submitted).
17. Joos, E. & Zeh, H. D. The emergence of classical properties through interaction with the environment. *Z. Phys. B* **59**, 223–243 (1985).
18. Kolodney, E., Budrevich, A. & Tsipinyuk, B. Unimolecular rate constants and cooling mechanisms of superhot C₆₀ molecules. *Phys. Rev. Lett.* **74**, 510–513 (1995).
19. Mitzner, R. & Campbell, E. E. B. Optical emission studies of laser desorbed C₆₀. *J. Chem. Phys.* **103**, 2445–2453 (1995).
20. Pfau, T., Spälter, S., Kurtsiefer, Ch., Ekstrom, C. R. & Mlynek, J. Loss of spatial coherence by a single spontaneous emission. *Phys. Rev. Lett.* **73**, 1223–1226 (1994).
21. Clauser, J. F. & Li, S. "Heisenberg microscope" decoherence atom interferometry. *Phys. Rev. A* **50**, 2430–2433 (1994).
22. Chapman, M. S. et al. Photon scattering from atoms in an atom interferometer: Coherence lost and regained. *Phys. Rev. Lett.* **75**, 3783–3787 (1995).
23. Werner, S. A., Colella, R., Overhauser, A. W. & Eagen, C. F. Observation of the phase shift of a neutron due to precession in a magnetic field. *Phys. Rev. Lett.* **35**, 1053–1055 (1975).
24. Rauch, H. et al. Verification of coherent spinor rotation of fermions. *Phys. Lett. A* **54**, 425–427 (1975).
25. Clauser, J. F. in *Experimental Metaphysics* (eds Cohen, R. S., Home, M. & Stachel, J.) 1–11 (Kluwer Academic, Dordrecht, 1997).
26. Arndt, M., Nairz, O., van der Zouw, G. & Zeilinger, A. in *Epistemological and Experimental Perspectives on Quantum Physics* (eds Greenberger, D., Reiter, W. L. & Zeilinger, A.) 221–224 (IVC Yearbook, Kluwer Academic, Dordrecht, 1999).

Acknowledgements

We thank M. Haluška, H. Kuzmany, R. Penrose, P. Scheier, J. Schmiedmayer and G. Senn for discussions. This work was supported by the Austrian Science Foundation FWF, the Austrian Academy of Sciences, the TMR programme of the European Union, and the US NSF.

Correspondence and requests for materials should be addressed to A.Z. (e-mail: zeilinger-office@exp.univie.ac.at).

Lanthanum-substituted bismuth titanate for use in non-volatile memories

B. H. Park*, B. S. Kang*, S. D. Bu*, T. W. Noh*, J. Lee† & W. Jo‡

* Department of Physics and Condensed Matter Research Institute, Seoul National University, Seoul 151-742, Korea

† Department of Materials Engineering, Sung Kyun Kwan University, Suwon 440-746, Korea

‡ LG Corporate Institute of Technology, Seoul 137-140, Korea

Non-volatile memory devices are so named because they retain information when power is interrupted; thus they are important computer components. In this context, there has been considerable recent interest^{1,2} in developing non-volatile memories that use ferroelectric thin films—'ferroelectric random access memories', or FRAMs—in which information is stored in the polarization state of the ferroelectric material. To realize a practical FRAM, the thin films should satisfy the following criteria: compatibility with existing dynamic random access memory technologies, large remnant polarization (P_r) and reliable polarization-cycling characteristics. Early work focused on lead zirconate titanate (PZT) but, when films of this material were grown on metal electrodes, they generally suffered from a reduction of P_r ('fatigue') with polarity switching. Strontium bismuth tantalate (SBT) and related oxides have been proposed to overcome the fatigue problem³, but such materials have other shortcomings, such as a high deposition temperature. Here we show that lanthanum-substituted bismuth titanate thin films provide a promising alternative for FRAM applications. The films are fatigue-free on metal electrodes, they can be deposited at temperatures of ~650 °C and their values of P_r are larger than those of the SBT films.

The structure of FRAM is very similar to that of conventional dynamic random access memory (DRAM), where memory cells are arranged in a square matrix. (Therefore, FRAM should, in principle, have a lower power requirement, a faster access time and a potentially lower cost than many other non-volatile memory devices^{1,3}.) A DRAM cell usually has a capacitor, where the binary information will be stored in terms of the signs of the stored charge. To maintain this information, a voltage should be continually applied to compensate for charge reduction due to leakage currents. In FRAM, the dielectric material in the capacitor is replaced with a ferroelectric film. Then, information can be stored in the polarization states of the ferroelectric thin film: that is, two spontaneous polarization states under zero electric field can be utilized as '0' and '1' digital states. It seems that the first generation of FRAM will be based on a destructive reading scheme, where a bit is read when a positive switching voltage (that is, an applied electric field) is applied to the memory cell². If the cell polarization is already along the same direction, then only a linear non-switching response, P_{ns} , is measured. If it is in the opposite direction, a switching response, P_{sw} , is measured. Therefore ($P_{sw} - P_{ns}$), which should be nearly the same as $2P_r$, is an important quantity which determines the performance of FRAM. It should be large enough for signal detection and should not change under repetitive read/write cycles.

The PZT and other related ferroelectric films, which are most widely investigated, usually have large values of ($P_{sw} - P_{ns}$): depending on substituting elements and processing conditions, reported values vary from 20 to 70 $\mu\text{C cm}^{-2}$ (refs 4, 5). However, when a capacitor is fabricated with the PZT film on conventional platinum (Pt) electrodes, its value of ($P_{sw} - P_{ns}$) is usually reduced after repetitive read/write cycles. This fatigue problem might be induced

by space charges resulting from (1) defects inside the ferroelectric material, (2) domain wall motion, or (3) defects near the electrode–ferroelectric interface³. In order to overcome the problem, modification of electrodes has been tried. When oxide or hybrid-metal-oxide electrodes are used instead of Pt electrodes, the fatigue problem in PZT films can be solved^{16–8}. However, these electrodes are more difficult to synthesize than pure metal electrodes such as Pt (ref. 9).

The use of SBT or related oxides was found to overcome the fatigue problem. They have the bismuth-containing layered perovskite structure where double layers of Ta–O octahedra are sandwiched between $(\text{Bi}_2\text{O}_2)^{2+}$ layers³. The $(\text{Bi}_2\text{O}_2)^{2+}$ layers should make SBT fatigue-free because the layers have net electrical charges, and so their positioning in the lattice is self-regulated to compensate for space charges. However, SBT films have relatively small values of $(P_{\text{sw}} - P_{\text{ns}})$, which vary from 4 to $16 \mu\text{C cm}^{-2}$ (refs 9–13). In addition, a high processing temperature (750–850 °C) makes them incompatible with conventional DRAM technologies.

Bismuth titanate, $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ (BTO), may be a good new candidate material for FRAM, since its bulk value of $2P_r$ is about $60 \mu\text{C cm}^{-2}$; it also has similar $(\text{Bi}_2\text{O}_2)^{2+}$ layers¹⁴. As shown in Fig. 1, it has triple layers of Ti–O octahedra between the $(\text{Bi}_2\text{O}_2)^{2+}$ layers. However, BTO films showed fatigue and unexpectedly low values of $(P_{\text{sw}} - P_{\text{ns}})$, 4–8 $\mu\text{C cm}^{-2}$ (refs 15, 16). The fatigue problem in the BTO films is interesting, as it suggests that the simple charge-compensating role of the $(\text{Bi}_2\text{O}_2)^{2+}$ layers is not sufficient to make the films fatigue-free.

Recently, using X-ray photoemission spectroscopy measurements, we investigated defects in SBT and BTO films¹⁷. For the SBT films annealed under an oxygen pressure of 10^{-4} torr, oxygen vacancies seemed to be located only near the $(\text{Bi}_2\text{O}_2)^{2+}$ layers. For BTO films annealed under similar conditions, oxygen vacancies were found at both the $(\text{Bi}_2\text{O}_2)^{2+}$ and the $(\text{Bi}_2\text{Ti}_3\text{O}_{10})^{2-}$ perovskite layers. Therefore, it appears that the chemical stability of the perovskite layers against oxygen vacancies may be another cause for the fatigue-free behaviour in the SBT films.

The oxygen ions near the Bi ions are likely to be less stable than those near Sr ions due to the volatility of the Bi ions. To test this idea,

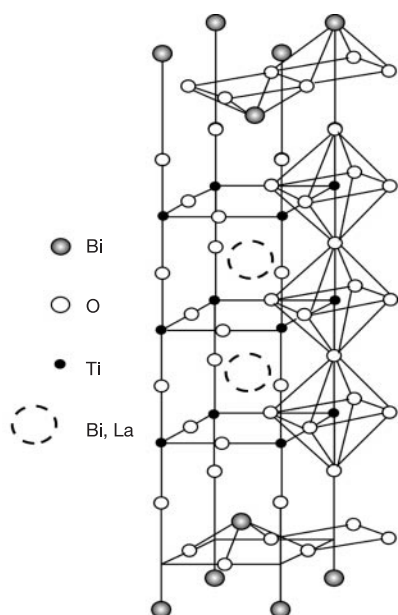


Figure 1 Lattice structure of the La-substituted bismuth titanate. The compound studied is $\text{Bi}_{3.25}\text{La}_{0.75}\text{Ti}_3\text{O}_{12}$ (BLT) where La ions can also be substituted by other rare-earth ions such as Pr, Nd, or Sm.

Table 1 Properties of PZT, SBT, and BLT thin films prepared on conventional Pt electrodes

	PZT	SBT	BLT
$P_{\text{sw}} - P_{\text{ns}}$ ($\mu\text{C cm}^{-2}$)	20–70 (refs 4, 5)	4–16 (refs 9–13)	16–20
Fatigue resistance	Poor	Good	Good
Leakage current at 5 V (A cm^{-2})	10^{-7} (ref. 3)	10^{-9} (ref. 3)	10^{-7}
Processing temperature (°C)	500–600 (refs 4, 9)	750–850 (refs 9–13)	650–700

PZT, lead zirconate titanate; SBT, strontium bismuth tantalate; BLT, $\text{Bi}_{3.25}\text{La}_{0.75}\text{Ti}_3\text{O}_{12}$.

we deposited $\text{Bi}_3\text{TiTaO}_9$ films, where all of the Sr ions in the SBT films are substituted with Bi ions and half of the Ta ions are substituted with Ti ions in order to obtain charge neutrality. Although the $\text{Bi}_3\text{TiTaO}_9$ films had electrical properties (such as $2P_r$ and retention characteristics) that were very similar to the SBT films, the $\text{Bi}_3\text{TiTaO}_9$ films showed fatigue failures, which suggests that the Bi ions do affect fatigue characteristics¹⁸. Hence, we deduced that the failure characteristics of BTO could be improved, if some Bi ions were substituted near the Ti–O octahedron layers with the La ions. We therefore investigated the electrical properties of $\text{Bi}_{3.25}\text{La}_{0.75}\text{Ti}_3\text{O}_{12}$ (BLT) films, whose bulk P_r value is the largest among those of La-substituted bismuth titanate materials. Also, it has a Curie point, which marks the transition between ferroelectric phase and paraelectric phase, of ~ 400 °C (ref. 19).

The BLT targets were prepared by solid-state reaction of a mixture of Bi_2O_3 , La_2O_3 and TiO_2 powders with a molar ratio of 13:3:24. Using pulsed laser deposition, we grew BLT films on Pt/Ti/SiO₂/Si substrates, which are commonly used for devices on Si substrates. The BLT films were deposited at 400 °C in 200 mtorr of oxygen for 5 min and post-annealed at 650 °C in an oxygen atmosphere for 1 h.

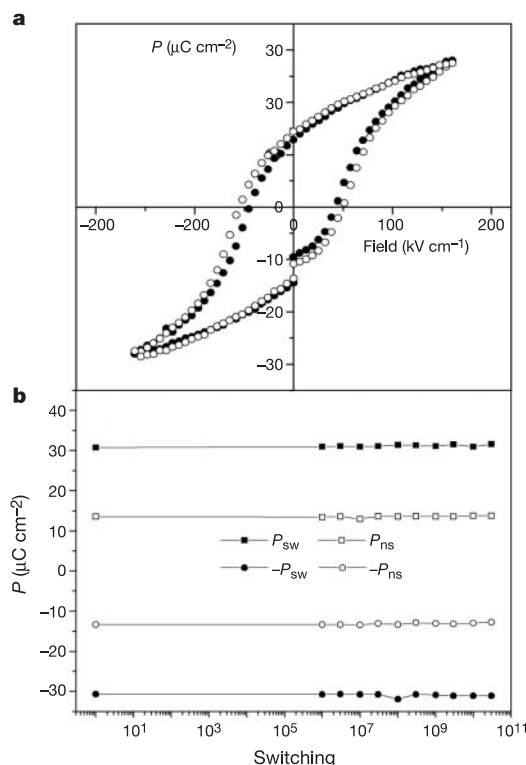


Figure 2 Results of the fatigue tests at 1 MHz. **a**, P – E hysteresis loops for the Au/BLT/Pt/Ti/SiO₂/Si films before (filled circles) and after (open circles) being subjected to 3×10^{10} read/write cycles. **b**, Variation of P_{sw} , P_{ns} , $-P_{\text{sw}}$, and $-P_{\text{ns}}$ versus number of cycles. Here $-P_{\text{sw}}$ and $-P_{\text{ns}}$ denote a switching polarization and linear non-switching polarization, respectively, when a negative read voltage is applied to the capacitor.

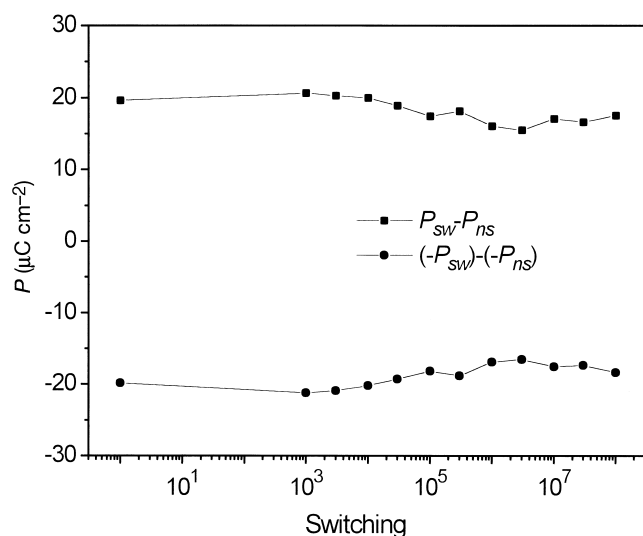


Figure 3 Results of the fatigue tests at 1 kHz. Shown is the variation of $(P_{sw} - P_{ns})$ and $(-P_{sw}) - (-P_{ns})$ with number of read/write cycles.

Note that the processing temperature is lower than that reported for SBT films by 100–200 °C. In order to fabricate a BLT capacitor structure, an Au top electrode was deposited at room temperature using the thermal evaporation method with a shadow mask. The area of the top electrode was $1 \times 10^{-4} \text{ cm}^2$.

Figure 2a shows typical well-saturated polarization–electric field (P – E) curves for a BLT capacitor before and after being subjected to 3×10^{10} read/write cycles at a frequency of 1 MHz. The P – E curves do not show any serious asymmetric behaviour resulting in imprint failures. The remnant polarization and the coercive field of the capacitor are about $12 \mu\text{C cm}^{-2}$ and 50 kV cm^{-1} , respectively. Figure 2b illustrates the change of polarizations after the capacitor has been subjected to various numbers of read/write cycles. At first, the values of P_{sw} and P_{ns} are about $32 \mu\text{C cm}^{-2}$ and $15 \mu\text{C cm}^{-2}$, respectively. (We note that the value of $(P_{sw} - P_{ns})$ is about four times as large as the values of SBT films deposited by pulsed laser deposition¹¹ or metal organic chemical vapour deposition¹². It is even larger than the largest reported value for SBT films synthesized by sol–gel methods^{3,13}.) Moreover, the value of $(P_{sw} - P_{ns})$ remains nearly constant after repetitive read/write cycles, and the shape of the P – E curve does not change significantly after 3×10^{10} cycles.

The fatigue characteristics at the low frequency of 1 kHz were also measured. As shown in Fig. 3, the values of $(P_{sw} - P_{ns})$ do not change significantly during the 1 kHz fatigue test. Note that the $(P_{sw} - P_{ns})$ values measured at 1 kHz are quite close to those measured at 1 MHz. It is clear that the BLT films have good fatigue resistance and large values of $(P_{sw} - P_{ns})$.

The dielectric constant and the dissipation factor of a BLT capacitor were also measured as a function of frequency. As shown in Fig. 4, the dielectric constant and the dissipation factor at 1 MHz are about 310 and 0.025, respectively. These values are comparable to those of PZT films and SBT films^{5,11}. The dielectric constant of the BLT capacitor slightly decreases with increasing frequency. There are no sudden changes of the dielectric constant and the dissipation factor within the frequency range up to 1 MHz. The typical leakage current density of the BLT films was found to be about $10^{-7} \text{ A cm}^{-2}$ at 250 kV cm^{-1} .

In Table 1, we summarize several important properties of PZT, SBT and BLT films. This table shows that the BLT films may be able to solve the fatigue problem observed in the PZT films. The BLT films have lower processing temperatures and larger values of $(P_{sw} - P_{ns})$ than the SBT films. There is some room for improve-

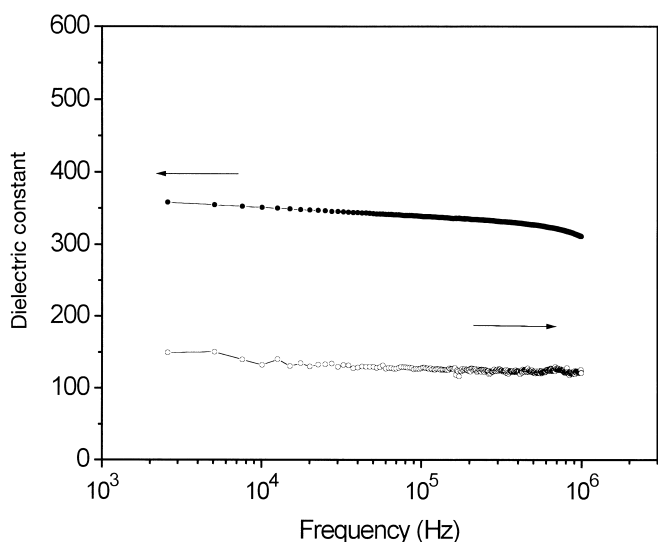


Figure 4 Dielectric constant and dissipation factor (loss) for the Au/BLT/Pt/Ti/SiO₂/Si films as a function of frequency.

ment by varying the amounts of the La substituents, or by substituting other rare-earth ions. □

Received 23 March; accepted 13 September 1999.

- Scott, J. F. & Araujo, C. A. Ferroelectric memories. *Science* **246**, 1400–1405 (1989).
- Auciello, O., Scott, J. F. & Ramesh, R. The physics of ferroelectric memories. *Phys. Today* **51**, 22–27 (1998).
- Araujo, C. A. *et al.* Fatigue-free ferroelectric capacitors with platinum electrodes. *Nature* **374**, 627–629 (1995).
- Keijser, M. & Dormans, G. J. M. Chemical vapor deposition of electroceramic thin films. *MRS Bull.* **21**, 37–43 (1996).
- Chen, H. D., Udayakumar, K. R., Gaskey, C. J. & Cross, L. E. Electrical properties' maxima in thin films of the lead zirconate–lead titanate solid solution system. *Appl. Phys. Lett.* **67**, 3411–3413 (1995).
- Ramesh, R. *et al.* Oriented ferroelectric La–Sr–Co–O/Pb–La–Zr–Ti–O/La–Sr–Co–O heterostructures on [001] Pt/SiO₂/Si substrates using a bismuth titanate template layer. *Appl. Phys. Lett.* **64**, 2511–2513 (1994).
- Nakamura, T., Nakao, Y., Kamisawa, A. & Takasu, H. Preparation of Pb(Zr,Ti)O₃ thin films on electrodes including IrO₂. *Appl. Phys. Lett.* **65**, 1522–1524 (1994).
- Al-Shareef, H. N., Bellur, K. R., Kingon, A. I. & Auciello, O. Influence of platinum interlayers on the electrical properties of RuO₃/Pb(Zr_{0.53}Ti_{0.47})O₃/RuO₂ capacitor heterostructures. *Appl. Phys. Lett.* **66**, 239–241 (1995).
- Auciello, O. & Ramesh, R. Laser-ablation deposition and characterization of ferroelectric capacitors for nonvolatile memories. *MRS Bull.* **21**, 31–36 (1996).
- Araujo, C. A. *et al.* Ferroelectric dielectric memory cell can switch at least giga cycles and has low fatigue—has high dielectric constant and low leakage current. *US Patent No.* 5,519,234 (1996).
- Dat, R., Lee, J. K., Auciello, O. & Kingon, A. I. Pulsed laser ablation synthesis and characterization of layered Pt/SrBi₂Ta₂O₉/Pt ferroelectric capacitors with practically no polarization fatigue. *Appl. Phys. Lett.* **67**, 572–574 (1995).
- Li, T. *et al.* Metalorganic chemical vapor deposition of ferroelectric SrBi₂Ta₂O₉ thin films. *Appl. Phys. Lett.* **68**, 616–618 (1996).
- Amanuma, K., Hase, T. & Miyasak, Y. Preparation and ferroelectric properties of SrBi₂Ta₂O₉ thin films. *Appl. Phys. Lett.* **66**, 221–223 (1995).
- Cummins, S. E. & Cross, L. E. Crystal symmetry, optical properties, and ferroelectric polarization of Bi₄Ti₃O₁₂ single crystals. *Appl. Phys. Lett.* **10**, 14–16 (1967).
- Joshi, P. C. & Krupanidhi, S. B. Switching, fatigue, and retention in ferroelectric Bi₄Ti₃O₁₂ thin films. *Appl. Phys. Lett.* **62**, 1928–1930 (1993).
- Kijima, T., Ushikubo, M. & Matsunaga, H. New low temperature processing of metalorganic chemical vapor deposition–Bi₄Ti₃O₁₂ thin films using Bi₂O₃ buffer layer. *J. Appl. Phys. (Jpn)* **38**, 127–130 (1999).
- Park, B. H. *et al.* Differences in nature of defects between SrBi₂Ta₂O₉ and Bi₄Ti₃O₁₂. *Appl. Phys. Lett.* **74**, 1907–1909 (1999).
- Kang, B. S. *et al.* Different fatigue behaviours of SrBi₂Ta₂O₉ and Bi₄Ti₃O₁₂ films: role of perovskite layers. *Appl. Phys. Lett.* (in the press).
- Takenaka, T. & Sakata, K. Electrical properties of grain-oriented ferroelectric ceramics in some lanthanum modified layer-structure oxides. *Ferroelectrics* **38**, 769–772 (1981).

Acknowledgements

This work was supported by the Korean Science and Engineering Foundation through the RCDAMP at Pusan National University, and by the Ministry of Science and Technology through the International Cooperative Research Project.

Correspondence and requests for materials should be addressed to T.W.N. (e-mail: twnoh@physa.snu.ac.kr).

Two-dimensional charge transport in self-organized, high-mobility conjugated polymers

H. Sirringhaus*, P. J. Brown*, R. H. Friend*, M. M. Nielsen†, K. Bechgaard‡, B. M. W. Langeveld-Voss‡, A. J. H. Spiering‡, R. A. J. Janssen‡, E. W. Meijer‡, P. Herwig§ & D. M. de Leeuw§

* Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

† Condensed Matter Physics and Chemistry Department, Risø National Laboratory, 4000 Roskilde, Denmark

‡ Laboratory of Macromolecular and Organic Chemistry, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands

§ Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Self-organization in many solution-processed, semiconducting conjugated polymers results in complex microstructures, in which ordered microcrystalline domains are embedded in an amorphous matrix¹. This has important consequences for electrical properties of these materials: charge transport is usually limited by the most difficult hopping processes and is therefore dominated by the disordered matrix, resulting in low charge-carrier mobilities² ($\leq 10^{-5} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$). Here we use thin-film, field-effect transistor structures to probe the transport properties of the ordered microcrystalline domains in the conjugated polymer poly(3-hexylthiophene), P3HT. Self-organization in P3HT results in a lamella structure with two-dimensional conjugated sheets formed by interchain stacking. We find that, depending on processing conditions, the lamellae can adopt two different orientations—parallel and normal to the substrate—the mobilities of which differ by more than a factor of 100, and can reach values as high as $0.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (refs 3, 4). Optical spectroscopy of the field-induced charge, combined with the mobility anisotropy, reveals the two-dimensional interchain character of the polaronic charge carriers, which exhibit lower relaxation energies than the corresponding radical cations on isolated one-dimensional chains. The possibility of achieving high mobilities via two-dimensional transport in self-organized conjugated lamellae is important for applications of polymer transistors in logic circuits⁵ and active-matrix displays^{4,6}.

We have studied the microstructure of 70–100 nm, spin-coated, regioregular P3HT films by grazing-incidence X-ray diffraction (XRD) on parts of the same SiO_2/Si substrates on which field-effect transistor (FET) devices were fabricated. Regioregularity denotes the percentage of stereoregular head-to-tail (HT) attachments of the hexyl side chains to the 3-position of the thiophene rings⁷. Two different orientations of the microcrystalline P3HT domains with respect to the FET substrate have been identified (Fig. 1). They are evident from the different intensity distributions of the (100) reflections due to the lamella layer structure and the (010) reflections due to π – π interchain stacking⁸. In samples with high regioregularity ($>91\%$) and low molecular weight the preferential orientation of ordered domains is with the (100)-axis normal to the film and the (010)-axis in the plane of the film (Fig. 1a). In contrast, in samples with low regioregularity (81%) and high molecular weight, the crystallites are preferentially oriented with the (100)-axis in the plane and the (010)-axis normal to the film (Fig. 1b)⁹. The cause of this surprising change of orientation is not fully understood. It must be a dynamic phenomenon during the rapid growth of spin-coated films affected by either regioregularity or molecular weight. In films prepared by slow casting from a dilute solution the (100)-axis is normal to the film for all polymers (green trace in Fig. 1c and d).

However, the ability to induce different orientations allows us to establish a direct correlation between the direction of π – π stacking and the in-plane FET mobility μ (Fig. 2a). At room temperature the highest mobilities of 0.05 – $0.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ are observed for the sample with the highest regioregularity (96%) and the largest size of crystallites with in-plane orientation of the (010)-axis ($\sim 95 \text{ \AA}$, as estimated from (010) line shape analysis¹⁰). For spin-coated samples with HT $\approx 81\%$, in which the (010)-axis is normal to the film, the mobility is only $2 \times 10^{-4} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ in spite of pronounced in-plane crystallinity along the (100)-axis with a grain size of $130 \text{ \AA}$. No clear correlation between FET mobility and weight average molecular weight M_w could be established, as seen previously between conductivity and M_w in highly doped P3HT (ref. 11). This is consistent with the FET mobility being limited by π – π interchain rather than intrachain transport as discussed below. In the case of the low-regioregularity polymers it is possible to directly compare mobilities for the in-plane (cast films) and out-of-plane orientation (spin-coated films) of the π – π stacking direction (Fig. 1d). In cast films of the 81% polymer the mobility is higher by more than an

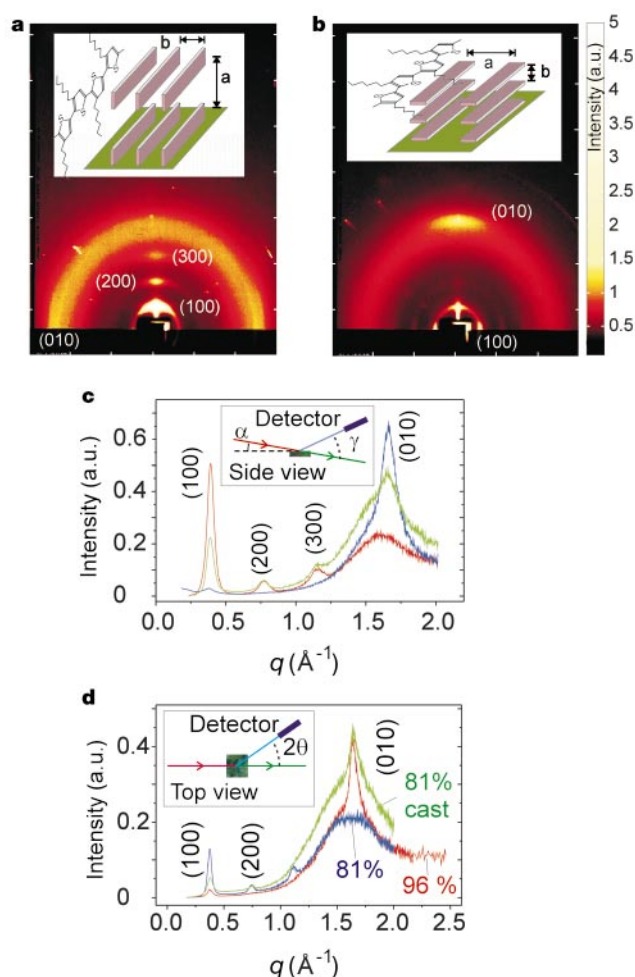


Figure 1 Two different orientations of ordered P3HT domains with respect to the FET substrate. **a, b**, The wide-angle X-ray scattering images are a colour representation of the two-dimensional distribution of scattered $\text{Cu K}\alpha$ X-ray intensity from spin-coated, 70–100 nm thick P3HT films with regioregularity of 96% (**a**) and 81% (**b**) on SiO_2/Si substrates. The vertical (horizontal) axes correspond to scattering normal (parallel) to the plane of the film. The insets show schematically the different orientations of the microcrystalline grains with respect to the substrate. **c, d**, The change of orientation is confirmed by high-resolution synchrotron X-ray diffraction measurements for constant, grazing-incidence angle with out-of-plane (**c**) and in-plane (**d**) scattering geometry. Red, 96%, spin coated; blue, 81%, spin coated; green, 81%, solution-cast. The intensities plotted versus the total scattering vector are corrected for polarization and geometric factors¹⁰.

order of magnitude than in spin-coated samples and only slightly lower than that of the highly regioregular polymers (Fig. 2a).

The large mobility anisotropy caused by different preferential orientations of the ordered, microcrystalline domains is clear evidence that the transport is no longer dominated by the remaining amorphous regions of the polymer film but is starting to reflect the transport properties of charge carriers in ordered polymer domains. The highest mobilities of $0.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and the mobility anisotropy for in- and out-of-plane π - π stacking of ~ 100 are of the same order of magnitude as in di-hexyl sexithiophene oligomer single crystals¹². The residual disorder in the films manifests itself in a thermally activated mobility at low temperatures (Fig. 2b). This is interpreted in terms of a distribution of disorder-induced, deeply localized states below the high-mobility electronic states at which charge transport occurs. The activation energies E_a extracted between room temperature and 150 K ($\mu \propto \exp(-E_a/kT)$) are comparable for the different polymers ($E_a = 84 \text{ meV}$ (96%), 100 meV (95%), 115 meV (81%)) suggesting similar degrees of disorder. From our experiments, the intrinsic mobility limit of a hypothetical P3HT single crystal cannot be estimated. However, we have entered a transport regime, in which at room temperature and high gate voltages the Fermi level at the interface is sufficiently close to the charge transport level that intrinsic transport effects such as mobility anisotropies reflecting local polymer self-organization can be observed.

We conclude that the high mobilities of $0.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ reported recently^{3,4} for the parallel orientation reflect efficient interchain transport in the two-dimensional conjugated lamellae. We discuss how these strong π - π interchain interactions may affect the nature of the charge-carrying species. Charge carriers in disordered polymers can be regarded as molecular radical cations formed on isolated, finite conjugated segments of the chain between two conjugation defects¹³. Due to polaronic relaxation of the local electronic and lattice structure the energy levels of a radical cation are shifted relative to those of the neutral molecule, giving rise to additional optical transitions upon charge injection (Fig. 3a). Charge modulation spectroscopy (CMS) measures changes of the optical transmission of a semitransparent FET upon gate-voltage induced modulation of the charge carrier density in the accumulation layer. It yields direct spectroscopic information about the charge carriers in polymer FETs (refs 14, 15) without the disturbing effects of counterions that are present in chemical doping experiments. The CMS spectra clearly show charge-induced, sub-gap transitions giving direct evidence for the polaronic nature of the

charge carriers. For 81% (96%) samples we observe a strong transition at 1.70 eV (1.75 eV), a second transition appearing as a shoulder at $1.35 \pm 0.05 \text{ eV}$ (Fig. 3a), and a third transition at 0.35 eV (0.32 eV) in the mid-infrared spectral range (Fig. 3b). The associated bleaching—that is, reduction of the strength of the π - π^* absorption ($\Delta T/T > 0$)—of those chains on which the charges are located (Fig. 3a) exhibits pronounced vibronic structure, which is more structured and red-shifted compared to the inhomogeneously broadened absorption spectrum of the film (P.J.B., H.S. and R.H.F., manuscript in preparation). This shows that the charge carriers observed in CMS have migrated to the most ordered domains in the film; this allows us to claim that the CMS spectra reported here yield the intrinsic spectroscopic signature of charge carriers in ordered P3HT domains. This is further corroborated by the observation that the CMS spectra in the visible and near-infrared exhibit no significant dependence of the relative intensities, energies, or the number of transitions on temperature T (100–300 K), modulation frequency, and carrier density (P.J.B., H.S. and R.H.F., manuscript in preparation). Therefore, the observed charge-induced transitions must belong to the same physical species, in contrast to CMS studies on oligothiophenes¹⁵ and less-ordered P3HT (ref. 14), in which there is coexistence of singly-charged radical cations, and doubly-charged dications and π -dimers. The observation of just one type of carrier in self-organized P3HT FETs is a manifestation of the low density of chemical and structural defects that may stabilize other carriers. From the independence of the CMS spectra with respect to temperature and charge carrier density we conclude that the carriers are singly-charged. Doubly-charged carriers would tend to dissociate at high temperatures and low carrier densities¹⁶.

In order to identify the spectroscopic signature of interchain interaction effects we compare the CMS spectra of P3HT FETs with chemical doping experiments on isolated oligo- and polythiophene radical cations (nT^+) in solution^{16,17}. For nT^+ two characteristic optical transitions, C1 and C2, are observed. Their energy red-shifts monotonically with increasing conjugation length (Fig. 3a). A third transition C3 is symmetry-forbidden on isolated chains¹⁶ and is not observed in the nT^+ spectra. The CMS spectra of self-organized P3HT presented here reveal characteristic features that cannot be understood as an extrapolation of the nT^+ radical cation spectra to long conjugation lengths: (1) We observe three transitions at 0.3–0.35 eV, 1.35 eV and 1.75 eV, the strongest of which (at 1.75 eV) is at significantly higher energy E than the charge-induced transitions of isolated nT^+ . It blue-shifts towards the π - π^* transition with increasing regioregularity ($\Delta E \approx 50 \text{ meV}$ between 81% and 96%)

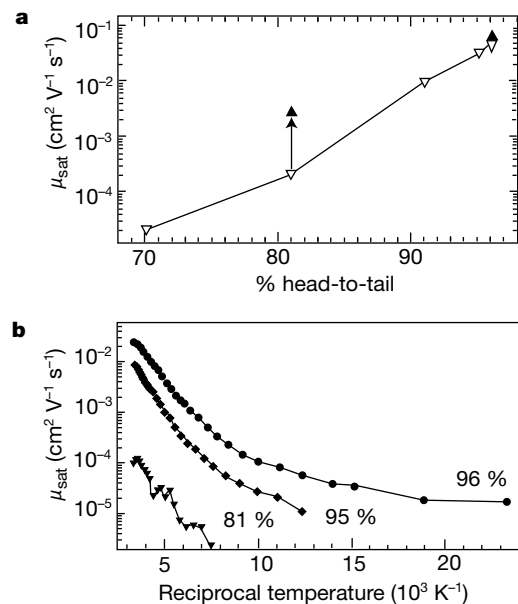


Figure 2 Charge carrier mobility of P3HT field-effect transistors with different microstructures. **a**, Dependence of the room-temperature mobility on the regioregularity for spin-coated (downward triangles) and solution-cast (upward triangles) top-contact P3HT FETs (channel length $L = 75 \mu\text{m}$, channel width $W = 1.5 \text{ mm}$). **b**, Temperature dependence of the field-effect mobility extracted from the characteristics of spin-coated bottom-contact P3HT FETs in the saturation regime ($L = 10 \mu\text{m}$, $W = 3 \text{ mm}$). Measurements were performed in vacuum ($p < 1 \times 10^{-6} \text{ mbar}$) to prevent charge trapping by adsorbed atmospheric impurities. The mobility of top-contact FETs with Au source-drain contacts evaporated after deposition of the polymer is higher, typically by a factor of two, than that of bottom-contact devices⁴.

whereas the infrared transition at ~ 0.35 eV red-shifts. (2) Below ~ 0.3 eV we observe low-energy transitions (CT) extending into the energy range of sharp charge-induced, vibrational 'amplitude modes' (AM)¹⁸. The intensity of the low-energy CT feature increases with regioregularity. Their overlap with the AM results in a sharp dip at 0.18 eV, which may be due to a Fano-type resonance caused by interactions of a discrete level with a continuum¹⁹. (3) The charge-induced transitions exhibit no vibronic structure in contrast to those of isolated oligomers, while at the same time the associated bleaching signal has a pronounced vibronic structure.

We regard these three characteristic features, not present in isolated radical cation spectra, as an experimental determination of the spectroscopic signature of the two-dimensional nature of charge carriers in the two-dimensional conjugated lamellae of ordered P3HT. In the presence of strong π - π interchain interactions, for which the microstructure-mobility correlation gives clear evidence, the carriers should no longer be confined to a single chain. Although the carriers still have a polaronic nature, requiring their wavefunctions to be localized, they have a pronounced interchain character, with wavefunctions spreading over neighbouring chains^{20,21}. This provides an explanation for the observed shift of

the energy levels towards the highest occupied (π) and lowest unoccupied (π^*) molecular orbitals (Fig. 3a), reflecting a lowering of the polaronic relaxation energy. The transitions at 0.35 eV, 1.35 eV and 1.75 eV may be related to the C1, C2 and C3 transitions, respectively, of the one-electron model of isolated radical-cations (Fig. 3a, $C1 + C2 \approx C3$). It has been suggested that, in the presence of interchain interactions, transitions such as C3 (that are symmetry-forbidden in isolated molecules) may become intense²¹. However, for a full theoretical understanding of the spectra a proper treatment of interchain and electron-electron interactions will be required. In the case of small relaxation energies and strong interchain interactions one may expect that due to configuration interactions¹⁶ charge-transfer like transitions (CT in Fig. 3) may become intense. They may explain the intense infrared absorption < 0.3 eV.

The evidence for an association of high-mobility with two-dimensional charge transport provides a firm basis for exploring supramolecular self-organization to enhance charge transport in conjugated polymer semiconductors. By better control of structural anisotropy, and by developing polymers with more strongly π - π interacting building blocks²², even higher mobilities and, possibly, a truly delocalized transport regime may be reached. □

Methods

Samples with HT regioregularity of 70% and 81% (molecular weight $M_w = 126$ kg mol⁻¹ and 175 kg mol⁻¹, respectively; polydispersity $D = 2.5$ –2.7) were synthesized by oxidative coupling with FeCl₃. Following the McCullough route, P3HT with regioregularity of 91% and 95% HT ($M_w = 11$ and 28 kg mol⁻¹, $D = 1.4$) was obtained. The highest regioregularity was a 96% HT sample synthesized by the Rieke route ($M_w = 28$ kg mol⁻¹, $D = 1.4$). For a review of the different synthesis routes see ref. 7. The regioregularity was determined by ¹H-NMR by comparing the signal intensities at 2.80 and 2.60 p.p.m. This yields a lower but more reliable value than analysis of the aromatic region of the spectrum. Films were spin-coated from a 0.8 weight % solution in CHCl₃ onto FET substrates (230 nm SiO₂ gate insulator on top of n⁺-Si gate electrodes) treated with the silylating agent hexamethyldisilazane to promote self-organization.

X-ray diffraction measurements were performed under inert He atmosphere to minimize air scattering and beam damage. A grazing incidence angle below the critical angle of total reflection from the substrate, but above the critical angle for the film, was chosen to enhance the sensitivity to the thin polymer film. Synchrotron XRD measurements were performed at the BW2 beamline of the German electron synchrotron facility (DESY) in Hamburg.

Received 12 March; accepted 3 August 1999.

- Samuelsen, E. J. & Mårdalen, J. In *Handbook of Organic Conductive Molecules and Polymers* Vol. 3 (eds Nalwa, H. S.) 87–120 (Wiley, Chichester, UK, 1997).
- Kobashi, M. & Takeuchi, H. Inhomogeneity of spin-coated and cast non-regioregular poly(3-hexylthiophene) films. Structures and electrical and photophysical properties. *Macromolecules* **31**, 7273–7278 (1998).
- Bao, Z., Dodabalapur, A. & Lovinger, A. J. Soluble and processable regioregular poly(3-hexylthiophene) for thin film field-effect transistor applications with high mobility. *Appl. Phys. Lett.* **69**, 4108–4110 (1996).
- Sirringhaus, H., Tessler, N. & Friend, R. H. Integrated optoelectronic devices based on conjugated polymers. *Science* **280**, 1741–1744 (1998).
- Drury, C. J., Mutsaers, C. M. J., Hart, C. M., Matters, M. & deLeeuw, D. M. Low-cost all-polymer integrated circuits. *Appl. Phys. Lett.* **73**, 108–110 (1998).
- Dodabalapur, A. et al. Organic smart pixels. *Appl. Phys. Lett.* **73**, 142–144 (1998).
- McCullough, R. D. The chemistry of conducting polythiophenes. *Adv. Mater.* **10**, 93–116 (1998).
- Prosa, T. J., Winokur, M. J., Moulton, J., Smith, P. & Heeger, A. J. X-ray structural studies of poly(3-alkylthiophenes)—An example of an inverse comb. *Macromolecules* **25**, 4364–4372 (1992).
- Fell, H. J., Samuelsen, E. J., Als-Nielsen, J., Grübel, G. & Mårdalen, J. Unexpected orientational effects in spin-cast, sub-micron layers of poly(alkylthiophene)s: A diffraction study with synchrotron radiation. *Solid State Commun.* **94**, 843–846 (1995).
- Warren, B. E. *X-ray Diffraction* 41–50 (Addison-Wesley, Reading, USA, 1969).
- Ishikawa, H. et al. Effect of molecular mass of poly(3-alkylthiophene) on electrical properties. *J. Phys. D* **25**, 897–900 (1992).
- Schön, J. H., Kloc, C., Laudise, R. A. & Batlogg, B. Electrical properties of single crystals of rigid rodlike conjugated molecules. *Phys. Rev. B* **58**, 12952–12957 (1998).
- Deussen, M. & Bässler, H. Anion and cation absorption spectra of conjugated oligomers and polymer. *Synth. Met.* **54**, 49–55 (1993).
- Ziemelis, K. E. et al. Optical spectroscopy of field-induced charge in poly(3-hexyl thiophene) metal-insulator-semiconductor structures: Evidence for polarons. *Phys. Rev. Lett.* **66**, 2231–2234 (1991).
- Harrison, M. G., Fichou, D., Garnier, F. & Yassar, A. In situ charge-modulation spectroscopy of oligothiophene field-effect diodes: from sexithiophene towards polythiophene. *Opt. Mater.* **9**, 53–58 (1998).
- Haare, J. A. E. H. V. et al. Redox states of long oligothiophenes: Two polarons on a single chain. *Chem. Eur. J.* **4**, 1509–1522 (1998).
- Horowitz, G., Yassar, A. & Bardeleben, H. J. V. ESR and optical spectroscopy evidence for a chain

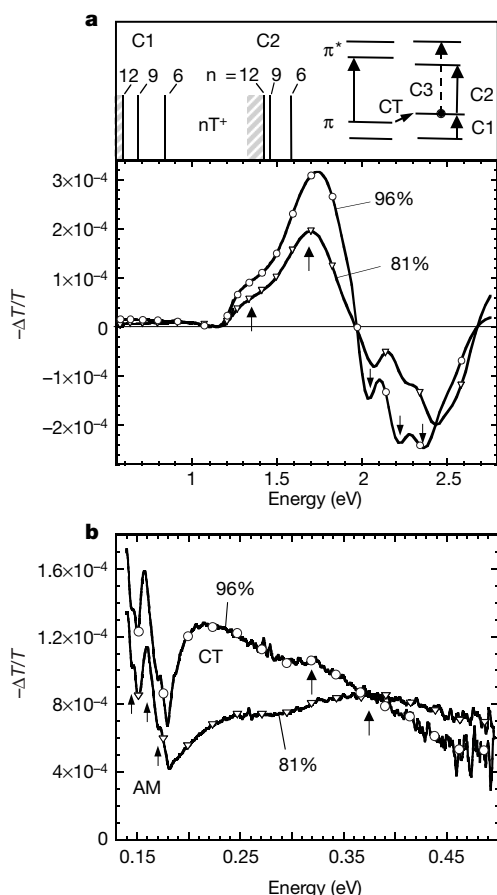


Figure 3 Charge modulation spectroscopy of semitransparent P3HT FETs in the accumulation regime. The devices were prepared under the same conditions and with similar mobilities as standard FETs and XRD samples (circles—96%, triangles—81%). **a, b**, The spectra show the fractional change of the optical transmission upon modulation of the gate voltage as a function of photon energy in the near-infrared and visible (**a**) and mid-infrared spectral range (**b**) ($T = 300$ K). Inset: schematic one-electron energy levels of neutral molecules and isolated radical cations, with the experimental energy value of the C1 and C2 transitions of nT^+ oligothiophene radical cations in solution (taken from ref. 16) shifting to lower energies with increasing n . In the solid state the transitions would be expected to be further red-shifted due to dielectric polarization effects. From the X-ray diffraction correlation lengths the conjugation length of the P3HT polymer is estimated to be $n \approx 20$ –25.

- length dependence of the charged states of thiophene oligomers. Extrapolation to polythiophene. *Synth. Met.* **62**, 245–252 (1994).
18. Horovitz, B. Infrared activity of Peierls systems and application to polyacetylene. *Solid State Commun.* **41**, 729–734 (1982).
19. Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124**, 1866–1878 (1961).
20. Emin, D. Self-trapping in quasi-one-dimensional solids. *Phys. Rev. B* **33**, 3973–3975 (1986).
21. Blackman, J. A. & Sabra, M. K. Interchain coupling and optical absorption in degenerate and nondegenerate polymers. *Phys. Rev. B* **47**, 15437–15448 (1993).
22. Sirringhaus, H. *et al.* Bis(dithienothiophene) organic field-effect transistors with a high ON/OFF ratio. *Appl. Phys. Lett.* **71**, 3871–3873 (1997).

Acknowledgements

We thank J. Cornil, D. Dos Santos, and A. J. Flewitt for contributions. Financial support from the European Commission and the Training and Mobility of Researchers (TMR) programme is acknowledged. H.S. thanks the Royal Society for a University Research Fellowship.

Correspondence and requests for materials should be addressed to H.S. (e-mail: hs220@phy.cam.ac.uk).

Identifying magma–water interaction from the surface features of ash particles

Ralf Büttner*, Pierfrancesco Dellino† & Bernd Zimanowski*

* Physikalisch Vulkanologisches Labor, Institut für Geologie, Universität Würzburg, Pleicherwall 1, D-97070 Würzburg, Germany

† Dipartimento Geomineralogico, Università degli Studi di Bari, Via Orabona 4, I-70125 Bari, Italy

The deposits from explosive volcanic eruptions (those eruptions that release mechanical energy over a short time span¹) are characterized by an abundance of volcanic ash^{2,3}. This ash is produced by fragmentation of the magma driving the eruption and by fragmenting and ejecting parts of the pre-existing crust (host rocks). Interactions between rising magma and the hydrosphere (oceans, lakes, and ground water) play an important role in explosive volcanism^{4,5}, because of the unique thermodynamic

properties of water that allow it to very effectively convert thermal into mechanical energy. Although the relative proportion of magma to host-rock fragments is well preserved in the pyroclastic rocks deposited by such eruptions, it has remained difficult to quantitatively assess the interaction of magma with liquid water from the analysis of pyroclastic deposits^{2–5}. Here we report the results of a study of natural pyroclastic sequences combined with scaled laboratory experiments. We find that surface features of ash grains can be used to identify the dynamic contact of magma with liquid water. The abundance of such ash grains can then be related to the water/magma mass ratios during their interaction.

From direct observation of ‘wet’ volcanic eruptions (phreatomagmatic eruptions) and the analysis of their deposits, several criteria have been defined that qualitatively indicate the presence of liquid water during an eruption^{2,3,6,7}. The reconstruction of volcanic eruptions from their deposits, however, would benefit from a tool that could quantitatively indicate the amount of liquid water that encountered the magma. A perfect laboratory volcano for investigations of magma–water contact is the active Italian volcanic island of Vulcano and its young crater, La Fossa, where various kinds of phreatomagmatic deposits, spanning the range of water–magma phenomenologies reported in the literature, are present.

Phreatomagmatic deposits at La Fossa di Vulcano have formed during almost all of the numerous eruptions characterizing its volcanic history, which started ~6,000 years ago, with the last eruption being the ‘vulcanian’ one of AD 1888–1890 (ref. 8). Following the scheme used in the literature^{9–11}, these deposits have been subdivided into two categories: ‘wet’ and ‘dry’ surge deposits, both being the products of laterally moving surges of low-density ash clouds. The difference between the two types is evidenced by their structural and textural features. ‘Dry’-type deposits are decimetres to metres thick densely laminated layers of coarse ash and lapilli; ‘wet’-type deposits are mostly not internally structured (that is, “massive”) centimetres thick fine-ash layers with features such as accretionary lapilli, vesiculated tuff, and plastic type deformation⁷.

Various structures of deposits have been interpreted using results from early magma–water interaction experiments¹². In ‘dry’ explosions the interaction is suggested to be more effective, with all water consumed during interaction and transformed into superheated steam, whereas in ‘wet’ explosions, the interaction is less effective,

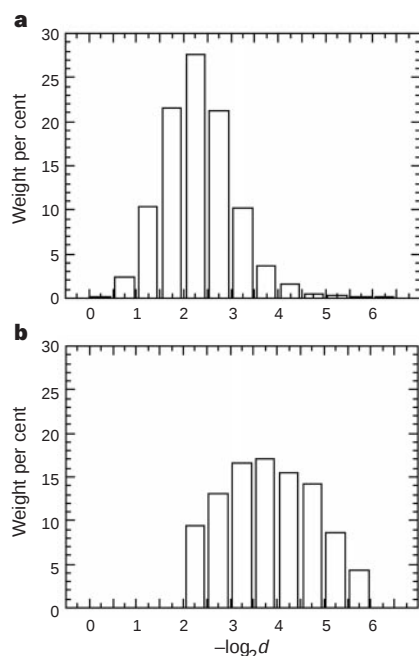


Figure 1 Grain size distribution of phreatomagmatic deposits of La Fossa di Vulcano. The figures show the proportional importance of grain size fractions in weight per cent (d is the sieve mesh size in millimetres). **a**, Histogram of a representative ‘dry’ surge deposit. **b**, Histogram of a representative ‘wet’ surge deposit.

with only some of the water involved (part of this is incorporated into the eruptive cloud as liquid droplets or condensing steam, which during the deposition leads to the features described above). In general, a higher amount of thermal energy is supposed to be transformed into mechanical energy during 'dry'-type eruptions.

This classical assumption may appear to be a good reference framework for the interpretation of sedimentary structures; however, the difference in thermal to mechanical energy conversion between the two types contrasts with the fact that 'wet' deposits, being much finer (Fig. 1), should be the result of more energetic fragmentation processes. We expected that what is recorded in the actual deposits mirrors a complex interplay between fragmentation and transportation mechanisms. Therefore, a detailed investigation of these deposits, and newly designed laboratory experiments, were undertaken in order to attain a precise interpretation.

The phreatomagmatic activity of La Fossa was experimentally simulated with the use of the TEE-II experiment¹³ at the Physikalisch Vulkanologisches Labor at Würzburg University. A suitable magma simulant was generated by remelting scoria from the La Sommata scoria cone, a small volcano in the vicinity of La Fossa. This material, following De Astis *et al.*¹⁴, closely resembles the original magmatic composition of the investigated La Fossa deposits. Three different experimental series were performed. In all cases 0.4 kg of melt at magmatic temperatures (1,570 K) were prepared under atmospheric pressure and controlled oxygen fugacity in cylindrical crucibles¹⁵. The temperature was raised ~100 K above the estimated temperature of the La Fossa magma, in order to establish a viscosity of 10 Pa s, which matches the viscosity of the original magma, as calculated from melt inclusion data¹⁶.

In two experimental series (series I and II) an optimum explosive

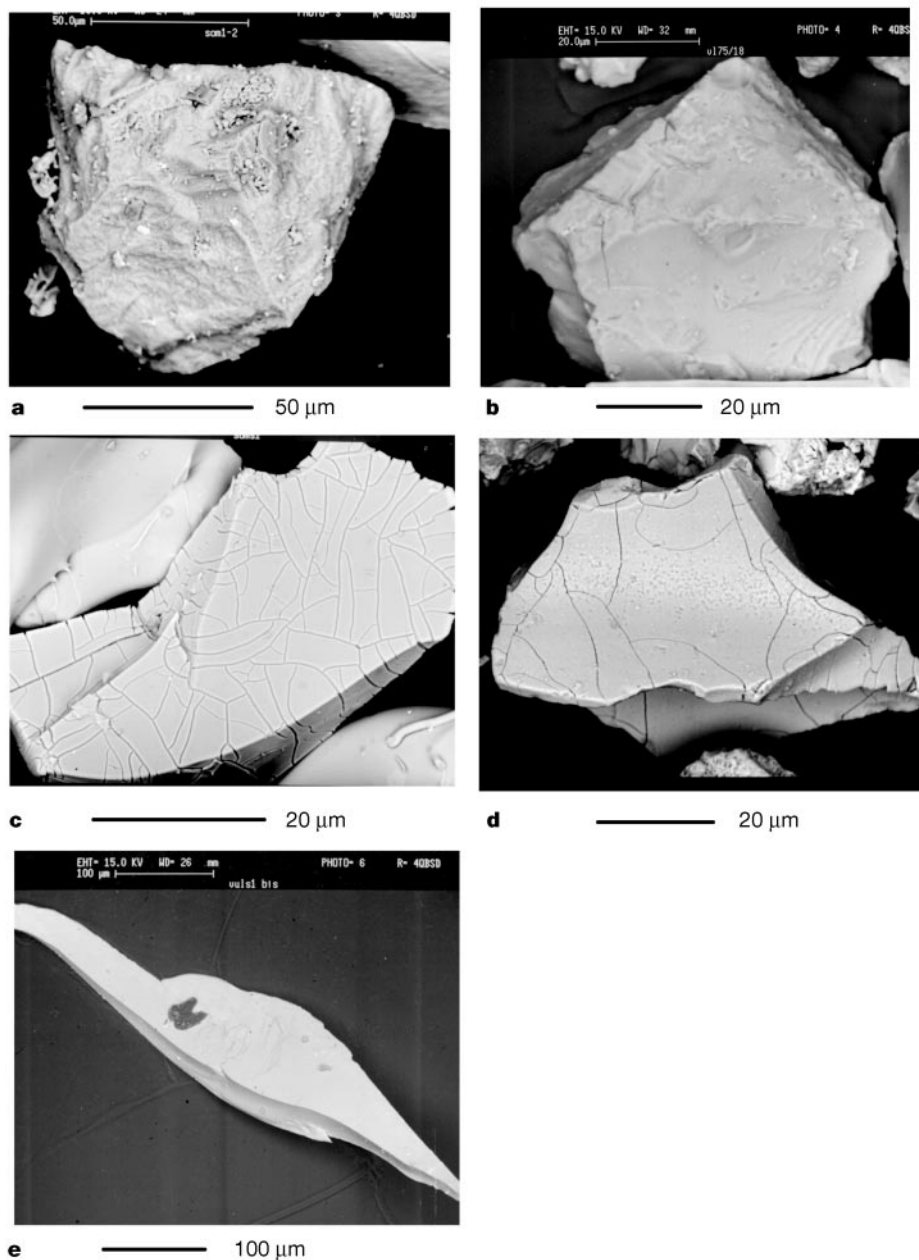


Figure 2 Comparison between characteristic particles from experiments with those from natural deposits of La Fossa di Vulcano using scanning electron microscope images.

a, Blocky and equant shaped particle from 'dry' explosion experiment (series I) with stepped surfaces and etches on the surface. **b**, Analogue particle from a 'dry' phreatomagmatic surge deposit of La Fossa. **c**, Particle of the same type as **a** and **b** from

'wet' explosion experiment (series II) showing additional branching quench-crack structures. **d**, Analogue particle from a 'wet' phreatomagmatic surge deposit of La Fossa. **e**, Elongated angular particle with planar surface textures from non-explosive thermal-granulation experiment (series III).

magma–water interaction was generated by injection of 10 ml of water at room temperature into the melt-filled crucibles¹⁵. In series I experiments, the particles formed by the explosion and consequent volumetric expansion of the contents of the crucibles in free air were recovered. This represented the ‘dry’ case (no excess water was present in the erupted system). In series II experiments, an additional water volume of 500 cm³ was introduced immediately before the onset of the explosion by flooding a cylindrical containment on top of the crucible in such a way that a water strata of 5 cm thickness was superimposed on the melt surface. Thus, the exploding system was forced to pass through the water strata before its final expansion in free air. This represented the ‘wet’ case (excess water was present in the erupted system). The resulting particles were also completely recovered.

In series III experiments, melt jets were produced by controlled tilting of the crucibles. These melt jets were poured into a water-filled container of 10 l volume from a height of about 30 cm. In this way non-explosive water–melt interaction was achieved, leading to thermal granulation of the melt. These particles were used as reference for ‘wet’ but non-explosive magma fragmentation. The experimentally produced ‘wet’ and ‘dry’ pyroclasts were then analysed and compared to their natural analogues.

In general, the artificial pyroclasts produced in the experiments cover all grain sizes and shape properties of the natural ones, thus illustrating the quality of the experimental scaling in terms of geometry and energy release. In contrast to the natural deposits, however, the mean sizes of the grain size distributions of the experimental particles are coarser.

The grain size distributions of series I (‘dry’) and series II (‘wet’) experimentally produced samples also do not mirror the granulometric differences found between actual wet and dry phreatomagmatic deposits of La Fossa (Fig. 1). In the experiments, the original post-explosive particle population is preserved. During a natural eruption, this population will be modified by post-explosive processes of eruption, transportation, and deposition, in fast-moving turbulent surge clouds. We conclude that these processes are primarily responsible for the granulometric features of the natural deposits.

It therefore seems more useful, when dealing with explosive fragmentation dynamics, to carefully examine the morphological features of natural and experimental ash particles. The significance of morphological micro-features on ash grains was predicted ~15 years ago¹², but the lack of analogue particles produced by scaled experiments at that time only allowed a qualitative discussion. The size fraction containing the relevant information on the explosive magma–water interaction processes is the fine-ash fraction¹⁷. Fine ash particles derived from ‘dry’ explosion experiments (series I) typically show blocky, equant shapes, and stepped glass surfaces by mechanical etching (Fig. 2a). The same type of particles are present in the fine ash fraction of natural ‘dry’ and ‘wet’ deposits (Fig. 2b). These particles clearly result from a brittle process, that is, the deformation rate acting on the magma was so high, that liquid relaxation of the stresses could not occur. These particles have been identified to be the fingerprints of thermohydraulic explosion processes¹⁸, representing the most intensive type of magma/water interaction.

Fine ash fragments from wet explosion experiments (series II) show similar features; in addition, on a significant number of fragments, the glass surfaces are covered by a branching network of cracks (Fig. 2c). These cracks must have formed immediately after fragmentation, as a result of sudden cooling. The formation of these patterns on the surfaces of experimentally produced particles by hydration processes can be excluded, because of the extremely short contact time between the hot particles and the water (<1 ms). Quench patterns of the same type are also found on corresponding particles from natural phreatomagmatic deposits (Fig. 2d), and in significantly higher amounts in wet surge deposits compared to dry

ones. Such quench patterns, that were never observed in the products of the dry experiments (series I), can thus be related to the fast passage (a few milliseconds) of newly fragmented particles through a domain of liquid water. This mechanism results in a high cooling rate and thus in an effective chilling. Particles of this kind only form when excess water is present in the conduit: they are more frequent in the ‘wet’ than in the ‘dry’ phreatomagmatic deposits of La Fossa. Only on the surfaces of the analogue natural particles, sometimes, skins of a few micrometres in thickness can be found, that may be related to hydration processes. The quench crack patterns, however, can be found on particles both with and without hydration skin. Thus, if hydration occurred, it affected the particles after the explosive fragmentation (syn-eruptive or post-eruptive).

The fragments produced by non-explosive thermal-granulation experiments (series III) typically show angular, elongated shapes, and planar glass surfaces (Fig. 2e). This kind of morphology cannot be found in the series I and II experiments and is also not represented in the natural dry or wet phreatomagmatic deposits. Therefore, non-explosive thermal granulation does not appear to play a major role in the magma–water interaction dynamics of La Fossa di Vulcano.

The quench crack structures that are exclusively observed on the surfaces of ash grains produced by explosive magma–water interaction under wet conditions are thus the witnesses of wet explosive phreatomagmatic volcanism, to be found in pyroclastic deposits of ancient and recent eruptions. The occurrence of a minor amount of these ‘marker’ particles in the ‘dry’ surge deposits of La Fossa thus indicates that, in this case, a moderate amount of excess water was present during the eruption. Future work on volcanic eruption mechanisms should include microscopic investigations of fine ash particles, rather than relying solely on grain size studies and the interpretation of macroscopic structural and textural features of pyroclastic deposits in the field. □

Received 6 May; accepted 23 August 1999.

- Freundt, A. & Rosi, M. (eds) *From Magma to Tephra* (Elsevier, Amsterdam, 1998).
- Fisher, R. V. & Schmincke, H.-U. *Pyroclastic Rocks* (Springer, Berlin, 1984).
- Heiken, G. & Wohletz, K. *Volcanic Ash* (University of California Press, Berkeley, 1985).
- Wohletz, K. H. Explosive magma–water interactions: thermodynamics, explosion mechanisms, and field studies. *Bull. Volcanol.* **48**, 245–264 (1986).
- Lorenz, V. Phreatomagmatism and its relevance. *Chem. Geol.* **62**, 149–156 (1987).
- Zimanowski, B. in *From Magma to Tephra* (eds Freundt, A. & Rosi, M.) 25–54 (Elsevier, Amsterdam, 1998).
- Lorenz, V. Vesiculated tuffs and associated features. *Sedimentology* **21**, 273–291 (1974).
- Frazzetta, G., La Volpe, L. & Sheridan, M. F. Evolution of the Fossa cone, Vulcano. *J. Volcanol. Geotherm. Res.* **17**, 329–360 (1983).
- Cas, R. A. F. & Wright, J. V. *Volcanic Successions* 528 (Unwin Hyman, London, 1987).
- Dellino, P., Frazzetta, G. & La Volpe, L. Wet surge deposits at La Fossa di Vulcano: depositional and eruptive mechanisms. *J. Volcanol. Geotherm. Res.* **43**, 215–233 (1990).
- De Rosa, R., Frazzetta, G. & La Volpe, L. An approach for investigating the depositional mechanism of fine-grained surge deposits. The example of the dry surge deposits at “La Fossa di Vulcano”. *J. Volcanol. Geotherm. Res.* **51**, 305–321 (1991).
- Wohletz, K. H. Mechanisms of hydrovolcanic pyroclast formation: grain-size, scanning electron microscopy, and experimental studies. *J. Volcanol. Geotherm. Res.* **17**, 31–64 (1983).
- Zimanowski, B., Büttner, R., Lorenz, V. & Häfele, H.-G. Fragmentation of basaltic melt in the course of explosive volcanism. *J. Geophys. Res.* **102**, 803–814 (1997).
- De Astis, G., La Volpe, L., Peccerillo, A. & Civetta, L. Volcanological and petrological evolution of Vulcano island (Aeolian Arc, southern Tyrrhenian Sea). *J. Geophys. Res.* **102**, 8021–8050 (1997).
- Zimanowski, B., Büttner, R. & Lorenz, V. Premixing of magma and water in MFCI experiments. *Bull. Volcanol.* **58**, 491–495 (1997).
- Gioncada, A. et al. A study of melt inclusions at Vulcano (Aeolian Islands, Italy): insights on the primitive magmas and on the volcanic feeding system. *Bull. Volcanol.* **60**, 286–306 (1998).
- Dellino, P. & La Volpe, L. Fragmentation versus transportation mechanisms in the pyroclastic sequence of Monte Pilato–Rocche Rosse (Lipari, Italy). *J. Volcanol. Geotherm. Res.* **64**, 211–232 (1995).
- Büttner, R. & Zimanowski, B. Physics of thermohydraulic explosions. *Phys. Rev. E* **57**, 5726–5729 (1998).

Acknowledgements

We thank K. Wohletz for his help in improving the quality of this Letter. This work was part of the EC Project “The physics of explosive volcanism”, supported by the European Community.

Correspondence and requests for materials should be addressed to B.Z. (e-mail: zimano@geologie.uni-wuerzburg.de).

Determinants of biodiversity regulate compositional stability of communities

Mahesh Sankaran & S. J. McNaughton

Biological Research Laboratories, Syracuse University, Syracuse, New York 13210-1244, USA

The world is witnessing a decline in biodiversity which may be greater in magnitude than even previous mass-extinction events^{1–3}. This has rekindled interest in the relationships between biodiversity and the stability of community and ecosystem processes⁴ that have been reported in some empirical studies^{5–7}. Diversity has been linked with community and ecosystem processes^{8–14}, but disputes remain over whether it is diversity, environmental factors or the variety of functional groups in a community that drive these patterns^{15–21}. Furthermore, it remains unclear whether variation in diversity resulting from species loss within communities has similar effects on stability as natural variation in diversity associated with gradients in factors that regulate diversity. We believe that, across larger ecological scales, extrinsic determinants of biodiversity such as disturbance regimes and site history may be the primary determinants of certain measures of community stability. Here we use controlled field experiments in savanna grasslands in southern India to demonstrate and explain how low-diversity plant communities can show greater compositional stability when subject to experimental perturbations characteristic of their native environments. These results are best explained by the ecological history and species characteristics of communities rather than by species diversity in itself.

We studied the responses of natural savanna-grassland communities to disturbance within the Kalakad-Mundanthurai Tiger Reserve (KMTR, 77° 15'–77° 40' E, 8° 25'–8° 55' N), along the southern section of India's Western Ghats Mountains. The stability of species composition of three low-elevation grassland types (200 m above sea level)—representing different positions along a productivity, diversity and disturbance gradient (see Methods)—were measured in 72 plots (each 4 m × 4 m) during 1997–98 following experimental perturbations in the form of fires, herbivore exclusion and simulated high-intensity grazing. The stability of species composition was characterized with two indices: (1) resistance to compositional change, R_c , measured as the change in the relative contribution of different species to the canopy between pre- and post-disturbance states²²; and (2) resistance to species turnover, R_{st} , calculated as the proportion of species common to pre- and post-disturbance plots.

Across all communities, compositional stability as measured by R_c was negatively correlated with diversity (Fig. 1a, using arcsin (R_c)^{0.5}: $r = -0.304$, $P = 0.009$), and low-diversity communities were more compositionally stable than high-diversity ones. In contrast, more diverse communities were more stable as measured by resistance to species turnover R_{st} (Fig. 1b, using arcsin (R_{st})^{0.5}: $r = 0.709$, $p < 0.001$).

As R_{st} is influenced both by patterns of species colonization and by loss from plots, each of these processes was analysed separately (Fig. 2a and b). The number of new species recorded in plots following the start of the experiment decreased as a function of initial diversity (Fig. 2a; $r = -0.482$, $P < 0.001$), while those lost from plots increased with diversity (Fig. 2b; $r = 0.617$, $P < 0.001$). In most cases, the number of colonizing species outweighed those lost from plots, resulting in the observed positive correlation

between diversity and R_{st} (Fig. 1b). For communities that share a common species pool, as was the case here, a negative relationship between diversity and colonization can result, even in the absence of specific ecological interactions, simply because fewer species remain in the pool to colonize species-rich plots. A positive correlation between species loss and diversity can result if high-diversity plots contain a greater number of rare species, which are likely to be lost due to purely stochastic processes. Across all plots, the number of rare species (cover < 1%) initially present was positively correlated with diversity ($r = 0.387$, $P < 0.05$). Irrespective of whether these observed trends were a consequence of an underlying ecological mechanism or a statistical phenomenon^{23,24}, low-diversity plots in this study had a greater turnover of species than high-diversity plots.

Unlike the relationships that appear so evident when data are pooled across communities, no consistent patterns were observed between diversity and either measure of stability within individual communities (arcsin (R_c)^{0.5}: *Cymbopogon flexuosus*: $r = 0.03$, $P > 0.05$; *Aristida setacea*: $r = -0.44$, $P < 0.05$; mixture: $r = -0.20$, $P > 0.05$; arcsin (R_{st})^{0.5}: *C. flexuosus*: $r = 0.65$, $P < 0.05$; *A. setacea*: $r = -0.14$, $P > 0.05$; mixture: $r = 0.29$, $P > 0.05$). These results do not constitute evidence for lack of significant effects of species diversity on the functioning of individual communities, as diversity was not specifically manipulated in these experiments. However, when coupled with the contrasting patterns observed between diversity and R_c and R_{st} across all

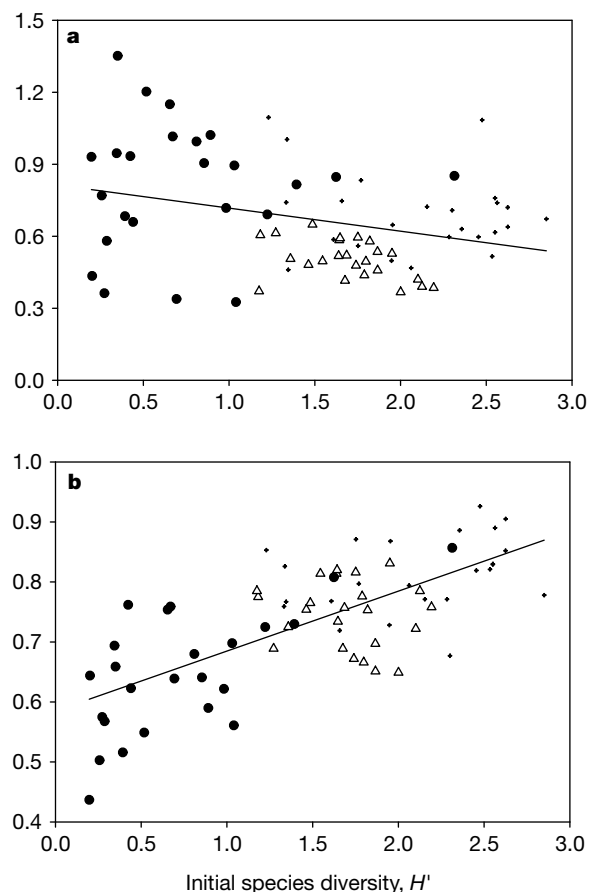


Figure 1 Compositional stability of communities as related to species diversity. **a**, Resistance to compositional change ($\arcsin(R_c)^{0.5} = -0.096H' + 0.813$) and **b**, resistance to species turnover ($\arcsin(R_{st})^{0.5} = 0.1H' + 0.585$), both plotted against initial species diversity (H') of 72 experimental plots. Symbols identify communities dominated by *C. flexuosus* (filled circles), *A. setacea* (open circles) and mixtures (dots).

communities (Fig. 1a and b), it does question the validity of absolute measures of species diversity, in and of themselves, as predictors of 'stability' in natural communities.

Species diversity in nature is an emergent property which results from historic, biotic and abiotic interactions among different constituent elements^{25,26}. Consequently, we may expect diversity in nature to co-vary with factors that regulate the distribution and abundance of species, such as disturbances, site productivity or site history (segregation of communities along the diversity axis in Fig. 1a and b). These factors influence the identities of potential member species in a community and can, therefore, affect its stability properties. To determine how much of the observed relationships in R_c and R_{st} were attributable to species and disturbance characteristics rather than diversity in itself, a multiple regression analysis was used on arcsin-transformed R_c and R_{st} to separate effects of community type, diversity, disturbance type and proneness to disturbance (see Methods). These variables cumulatively explained 42% of the observed variation in R_c (multiple $R^2 = 0.424$; $P < 0.001$), but only community type and proneness to disturbance had significant effects ($P < 0.001$). On the other hand, 53% of the variation in R_{st} was explained by variables included in the regression (multiple $R^2 = 0.529$; $P < 0.01$), but only diversity was significant in this case (H' : $P < 0.001$).

Greater species turnover in low-diversity communities in this study did not translate into lowered stability as measured by R_c as most species that colonized (or were lost from) plots were rare and did not contribute significantly to total cover. Compositional stability as measured by R_c depends on the sum total of shifts in

relative cover of individual species. Dominant species are likely to have disproportionate effects on R_c as they are capable of larger absolute shifts in cover compared to rare species. Low-diversity systems dominated by one or a few species can therefore show a large range of variation in compositional stability, depending on the response of the dominant species (triangular scatter of data points in Fig. 1a). Similar patterns of greater variation in community properties such as CO_2 flux¹³, biomass and density¹⁴ at lower diversities have also been reported for synthesized microbial communities. Even though low-diversity communities may show greater variation in levels of stability, they may be more stable than some higher-diversity communities when the dominant species responds 'favourably' to the disturbances in question. Given the role that disturbances play in structuring natural communities, such patterns may be more common in nature than is currently believed.

Previous studies investigating the biodiversity–stability relationship have focused on aggregate community properties such as biomass, productivity and nutrient cycling, while the relationship between diversity and constancy of species composition has received less attention²². Greater stability of aggregate community properties with increasing diversity has been argued to result from the increased probability of species or functional groups being present that can adequately compensate for those harmed by the disturbance^{6,8,11,14,27}. However, compensation, by definition, implies compositional change. As our data show, species and dominance characteristics (collectively identified by community type in this study) and disturbance history (as indexed by proneness to disturbance) may better explain compositional stability patterns across different community types.

These results have several implications for community, restoration and conservation biologists. First, it is critical that different aspects of the biodiversity–stability relationship arising from different choices of spatial and temporal scales not be confused. Evidence for a negative effect of lowered species diversity on stability resulting from species loss in a community in ecological time does not imply that species-poor communities in nature, which have evolved over evolutionary time, are necessarily less stable than species-rich ones. Second, as our data indicate, and as has been noted previously²⁸, patterns of community stability vary depending on the specific process measured. Third, disturbance regimes, site productivity and other environmental factors that are currently being modified can alter the stability properties of communities. *In situ* declines in species diversity due to local extinctions can further modify these patterns. Last, evidence for stability of aggregate community properties such as nutrient cycling or above-ground biomass does not preclude compositional shifts in communities^{6,7}. Communities and ecosystems are more than the biomass that they support or the nutrients that they cycle. Even though biomass or rates of nutrient cycling may remain unchanged, altered abundance of food or host plants can change, and potentially destabilize, herbivore and dependent predator populations. Although high biodiversity may in some cases be associated with 'desirable' responses such as stability of nutrient cycling or productivity, we warn against concluding that species-rich ecosystems will necessarily 'cope' better than species-poor ones in the face of perturbations. □

Methods

Communities

Responses of three different community-types to disturbances were studied at KMTR during 1997–98. Communities dominated by *C. flexuosus* were the most productive, but had low species richness and sustained low levels of herbivory. *A. setacea*-dominated communities were the least productive, had intermediate species richness and suffered highest levels of herbivory. Communities that had both species present were intermediate in productivity and levels of herbivory sustained, and had the highest species richness. Communities also differed in fire-proneness, with *C. flexuosus* communities the most fire-prone, and *A. setacea* communities the least. The contribution of grass species to the

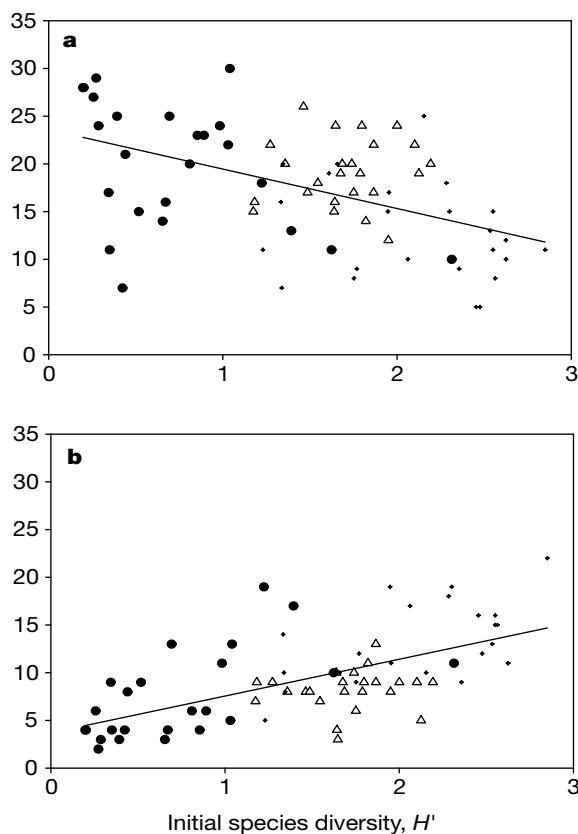


Figure 2 Patterns of species colonization and loss from plots. **a**, The number of new species that colonized plots ($NEW = -4.123H' + 23.58$) and **b**, the number of species that were lost from plots ($LOST = 3.856H' + 3.692$) after 1 year, both plotted against initial species diversity (H') of experimental plots. Symbols as in Fig. 1.

understory canopy varied from 85% in *C. flexuosus* communities, to 70% in *A. setacea* communities, and 60% in mixture communities.

Experimental design

The experimental design was a 2 × 3 factorial experiment with two burning treatments (burned and unburned) and three grazing treatments (natural levels of grazing, ungrazed and experimentally clipped). All experimental plots were 4 m × 4 m, located within an area of ~1 km², subject to similar climate conditions and potentially shared a common species pool. Overall, there were nine replicates for each unburned treatment and 18 for treatment combinations involving burning (three and six in each of three community types, respectively). Plots not experimentally manipulated (unburned, naturally grazed treatments) were excluded from the analysis. At each sampling session, species richness and cover was enumerated in eight and four 1-m² sub-plots, respectively, using a stratified sampling scheme. Species cover in sub-plots was estimated using a 1 m × 1 m grid frame subdivided into 100 units of 0.01 m² each. Data reported here are for one year following experimental manipulations, and are therefore devoid of any seasonal biases. Where necessary, they were transformed to fit the assumptions of normality.

Indices

$R_c = \sum \text{minimum} (p_{hi}, p_{it})$, where p_{hi} and p_{it} represent the relative cover of the i th species in pre-disturbance and 1 year post-disturbance plots, respectively²³. $R_q = N_{com}/N_{tot}$, where N_{tot} represents the total number of distinct species recorded in pre-disturbance and 1 year post-disturbance plots, and N_{com} represents the number of species common to pre-disturbance and 1 year post-disturbance plots. Diversity was calculated using the Shannon–Weiner index²⁹ as $H' = \sum p_i \ln(p_i)$, where p_i represents the proportional contribution of the i th species to the canopy. Proneness of communities to different disturbance combinations was calculated as $P_{bg} = P_b + P_g$, where the subscripts b and g represent the specific burning and grazing treatments. Proneness to burning was determined on the basis of the cover of *C. flexuosus* present initially in the plot (P_c). Burned treatments were assigned the value P_c whereas unburned treatments were assigned a value of $(1 - P_c)$ for this index. We believe that this is a valid index because *C. flexuosus* individuals are characteristic of the fire-prone environments, and also promote fires because of the extent of litter and standing dead biomass they produce. For the grazing treatments, grazed and clipped plots were assigned the value P_g and ungrazed plots $(1 - P_g)$, where P_g represents the fraction of species initially grazed in plots.

Received 27 July; accepted 16 August 1999.

1. Wilson, E. O. *The Diversity of Life* (Norton, New York, 1992).
2. Steadman, D. W. in *Biodiversity II* (eds Reaka-Kudla, M. L., Wilson, D. E. & Wilson, E. O.) 139–162 (Joseph Henry, Washington DC, 1997).
3. Russell, G. J., Brooks, T. M., McKinney, M. M. & Anderson, C. G. Present and future taxonomic selectivity in bird and mammal extinctions. *Conservation Biol.* **12**, 1365–1376 (1998).
4. Schulze, E. D. & Mooney, H. A. (eds) *Biodiversity and Ecosystem Function* (Springer, Berlin, 1993).
5. Hurd, L. E., Mellinger, M. V., Wolf, L. L. & McNaughton, S. J. Stability and diversity at three trophic levels in terrestrial ecosystems. *Science* **173**, 1134–1136 (1971).
6. McNaughton, S. J. Diversity and stability of ecological communities: a comment on the role of empiricism in ecology. *Am. Nat.* **111**, 515–525 (1977).
7. McNaughton, S. J. in *Biodiversity and Ecosystem Function* (eds Schulze, E. D. & Mooney, H. A.) 361–383 (Springer, Berlin, 1993).
8. Tilman, D. & Downing, J. A. Biodiversity and stability in grasslands. *Nature* **367**, 363–365 (1994).
9. Naem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining diversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
10. Naem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Empirical evidence that declining diversity can alter the performance of ecosystems. *Phil. Trans. R. Soc. Lond. B.* **347**, 249–262 (1995).
11. Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
12. Naem, S., Hakansson, K., Lawton, J. H., Crawley, M. J. & Thompson, L. J. Biodiversity and productivity in a model assemblage of plant species. *Oikos* **76**, 259–264 (1996).
13. Grady-Steed, J., Harris, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).
14. Naem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).
15. Tilman, D. *et al.* The influence of functional diversity on ecosystem processes. *Science* **277**, 1300–1302 (1997).
16. Hooper, D. U. & Vitousek, P. M. The effect of plant composition and diversity on ecosystem processes. *Science* **277**, 1302–1305 (1997).
17. Hooper, D. U. & Vitousek, P. M. Effects of plant composition and diversity on nutrient cycling. *Ecol. Monogr.* **68**, 121–149 (1998).
18. McGillivray, C. W. *et al.* Testing predictions of the resistance and resilience of vegetation subjected to extreme events. *Functional Ecol.* **9**, 640–649 (1995).
19. Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
20. Grime, J. P. Biodiversity and ecosystem function: the debate deepens. *Science* **277**, 1260–1261 (1997).
21. Wardle, D. A. A more reliable design for biodiversity study? *Nature* **394**, 30 (1998).
22. Frank, D. A. & McNaughton, S. J. Stability increases with diversity in plant communities: empirical evidence from the 1988 Yellowstone drought. *Oikos* **62**, 360–362 (1991).
23. Doak, D. F. *et al.* The statistical inevitability of stability–diversity relations in community ecology. *Am. Nat.* **151**, 264–276 (1998).
24. Tilman, D., Lehman, C. L. & Bristow, C. E. Diversity–stability relationships: statistical inevitability or ecological consequence. *Am. Nat.* **151**, 277–282 (1998).
25. Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).

26. Barbault, R. & Sastrapradja, S. in *Global Biodiversity Assessment* (ed. Heywood, V. H.) 193–274 (Cambridge Univ. Press, 1995).
27. Tilman, D. Biodiversity: population versus ecosystem stability. *Ecology* **77**, 350–363 (1996).
28. Pimm, S. L. *The Balance of Nature?* (Univ. Chicago Press, 1991).
29. MacArthur, R. H. Fluctuations of animal populations and a measure of community stability. *Ecology* **36**, 533–536 (1955).

Acknowledgements

We thank the Tamil Nadu Forest Department for granting permission to work at KMTR, and J. Ratnam for support and comments. We also thank L. L. Wolf, D. Frank, T. R. Shankar Raman and D. Barua for comments; R. Ali, V. Vinatha, K. Kar Gupta, M. Katti, D. Mudappa, N. M. Ishwar, K. Vasudevan and K. S. Gopi for their help; and C. Sankaran, P. Kumar and C. Jayseelan for field assistance. This work was supported by the Wildlife Conservation Society (India), NSF and the Sophie Danforth Conservation Biology Fund.

Correspondence and requests for materials should be addressed to M.S. (e-mail: msankara@mailbox.syr.edu).

Symmetry in locomotor central pattern generators and animal gaits

Martin Golubitsky*, Ian Stewart†, Pietro-Luciano Buono‡ & J. J. Collins§

* Mathematics Department, University of Houston, Houston, Texas 77204-3476, USA

† Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

‡ Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK

§ Department of Biomedical Engineering, Center for BioDynamics, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA

Animal locomotion is controlled, in part, by a central pattern generator (CPG), which is an intraspinal network of neurons capable of generating a rhythmic output^{1–4}. The spatio-temporal symmetries of the quadrupedal gaits walk, trot and pace^{5–8} lead to plausible assumptions about the symmetries of locomotor CPGs^{9–11}. These assumptions imply that the CPG of a quadruped should consist of eight nominally identical subcircuits, arranged in an essentially unique matter. Here we apply analogous arguments to myriapod CPGs. Analyses based on symmetry applied to these networks lead to testable predictions, including a distinction between primary and secondary gaits, the existence of a new primary gait called ‘jump’, and the occurrence of half-integer wave numbers in myriapod gaits. For bipeds, our analysis also predicts two gaits with the out-of-phase symmetry of the walk and two gaits with the in-phase symmetry of the hop. We present data that support each of these predictions. This work suggests that symmetry can be used to infer a plausible class of CPG network architectures from observed patterns of animal gaits.

The architecture of CPGs is seldom observable *in vivo*. Aspects of CPG structure are therefore usually inferred from observable gait features such as the phase of the gait cycle at which a given limb hits the ground, and the ‘duty factor’—the proportion of the gait cycle that a limb is in contact with the ground. It is usual to model CPGs as networks of nominally identical systems of differential equations, variously described^{9–17} as ‘units’, ‘oscillators’ or ‘cells’. We use the term ‘cell’.

Here we discuss a schematic CPG network¹⁰ (Fig. 1) that has twice as many cells as the animal has legs. For expository purpose we assume that cells 1, ..., 2*n* determine the timing of leg movements, and refer to the remaining cells as ‘hidden’.

The structure of the CPG network for a quadruped shown in Fig. 1b can be deduced from six assumptions: (1) the abstract CPG network is composed of identical cells, and the signal from each cell goes to one leg; (2) different gaits are generated by the same

network, and switching between gaits arises from changes in parameters (such as coupling strengths); (3) the locomotor CPG has the same architecture for all quadrupeds; (4) the network can generate the rhythms of walk, trot and pace; (5) trot and pace are dynamically independent, existing in different regions of parameter space; (6) the network generates only simple rhythmic patterns observed in quadrupedal locomotion.

The first assumption introduces symmetry into model CPGs^{9–11,15} and accounts for the observed symmetries of many gaits^{5–8}. Different CPGs could control the rhythms of each gait; but there would then have to be a controller that activates each CPG at the correct times, and there is no evidence for such controllers. It seems more likely that the second assumption is valid. Similarly, there is an evolutionary advantage in having a single architecture valid for all quadruped locomotor CPGs, thus justifying the third assumption.

As virtually all quadrupeds walk and either trot or pace, the network must produce these gaits, and so the fourth assumption seems appropriate. As camels pace but do not trot, and squirrels trot but do not pace, the fifth assumption must be valid. Finally, the sixth assumption is required in order to create the simplest CPG model.

We prove¹⁸ that any CPG satisfying all six assumed conditions must consist of eight identical cells, whose interconnections have the same symmetry as Fig. 1b. There are two types of symmetry in the network: contralateral symmetry, that interchanges cells on the left with cells on the right; and ipsilateral symmetry, that cyclically and simultaneously permutes cells on both left and right.

There are two types of symmetries of a periodic solution: spatial symmetries, which fix the solution at each point in time; and spatio-temporal symmetries, which fix the solution only after a phase shift. For example, in a pace, interchanging fore and hind legs is a spatial symmetry, whereas interchanging right and left legs coupled with a half-period phase shift is a spatio-temporal symmetry.

Symmetries of differential equations force well-defined spatio-temporal relations (like half-period phase shifts) on periodic solutions¹⁹. A symmetry type of periodic solutions is called robust if small symmetric perturbations of the equations do not change the symmetries of the periodic solution. If no symmetries are present, then small perturbations of the equations are extremely likely to change the phase relations.

For coupled cell systems, we prove the following¹⁸. Let H be the subgroup of spatio-temporal symmetries and let $K \subset H$ be the subgroup of spatial symmetries; then the only robust types of periodic solutions of coupled cell systems are those for which the

Table 1 Phase shifts for primary gaits in the eight-cell network

		Walk	Jump	Trot	Pace	Bound	Pronk
LF	RF	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$
LH	RH	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$
LF	RF	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$
LH	RH	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$	$\pm \frac{1}{2}$

LF, left fore leg; RF, right fore leg; LH, left hind leg; RH, right hind leg.

quotient group H/K is cyclic. The most symmetric solutions are the ones for which the group H is the symmetry group of the network.

Table 1 lists the most symmetric rhythms; these correspond to standard quadrupedal gaits with the exception of the 'jump'. For concrete networks, the values of the coupling terms in the network dynamics determine which rhythms occur (refs 18, 20; P.-L. Buono, manuscript in preparation).

The eight-cell network in Fig. 1b is the only one that satisfies all six assumptions^{10,18}. We sketch the argument. Walk implies that the network must have a four-cycle symmetry, and either trot or pace implies that the network must have a two-cycle symmetry. These statements follow from the first and fourth assumptions. The second assumption implies that the network must possess both of these symmetries.

The four-cycle symmetry implies that the number of cells is a multiple of four. If the network has four cells, then the two-cycle and four-cycle symmetries cannot commute and this forces trot and pace to be conjugate solutions, contradicting the fifth assumption. If the network has more than eight cells, then unnatural phase-shift symmetries would be allowed, negating the sixth assumption. Thus there are eight cells. In an eight-cell network, the fifth assumption implies that the four-cycle and two-cycle symmetries commute, and this leads to the network whose symmetry type is indicated in Fig. 1b. By analogy, the gaits of $2n$ -legged animals require a network with the same symmetries as Fig. 1c.

Even though our model prescribes only the symmetries of the CPG, and not the differential equations, it still leads to several predictions:

Prediction 1. In our model, a gait is primary if the signal from each cell is identical up to phase shift. Secondary gaits are those where the signals to different legs are different waveforms. Our network produces six primary gaits¹⁰ (see Table 1) and solutions that resemble gallops and canters (refs 10, 20; P.-L. Buono, manuscript

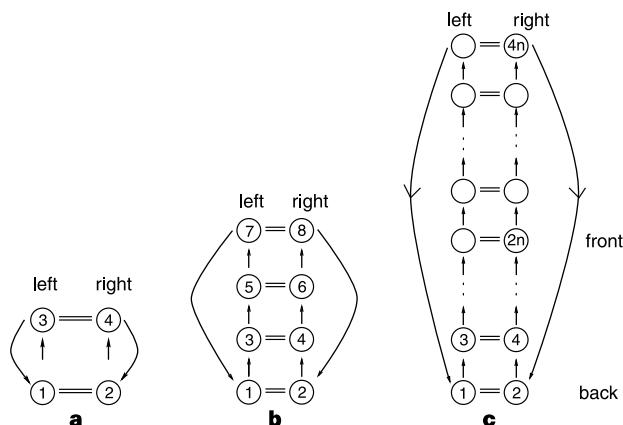


Figure 1 Schematic central pattern generator (CPG) networks. **a**, Four-cell network for bipedal locomotor CPG; **b**, eight-cell network for quadrupeds; **c**, $4n$ -cell network for $2n$ -legged animals. Double lines indicate contralateral coupling; single lines indicate ipsilateral coupling. Direction of ipsilateral coupling is indicated by arrows; contralateral coupling is bidirectional.

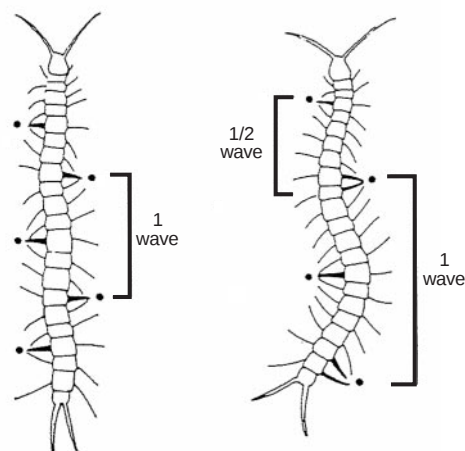


Figure 2 Half-integer wave numbers in myriapod gaits. Figure of Manton²⁵ from Alexander⁵. Thick lines indicate centipede legs in contact with the ground. Left centipede shows 3 waves. Right centipede shows $3/2$ waves.

Table 2 Bipedal gaits for a four-cell CPG network

Slow hop		Fast hop		Walk		Run	
0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$\frac{1}{2}$
0	0	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$

in preparation), which are secondary gaits. Experimentally observed variability from exact quarter-period and half-period phase shifts^{21,22} can, in principle, be accounted for by analysing symmetry-broken secondary gaits. As the signals sent to fore legs in primary gaits are identical and in secondary gaits are unequal, the duty factors of the fore legs should be equal in primary gaits and unequal in secondary gaits. Indeed, the duty factors of fore legs of a walking horse are equal⁸ and those of a galloping horse are different^{23,24}.

Prediction 2. Table 1 includes a non-standard primary gait, the jump, which can be described as 'fore feet hit ground, then hind feet hit ground, then three beats later fore feet hit ground'. We observed a gait with that pattern in a bucking bronco. (A figure showing four equal-time-interval frames, taken from a video provided by The Houston Livestock Show and Rodeo of the bareback bronco event on 24 February 1999, is available electronically at <ftp://ftp.math.uh.edu/pub/laode/rodeo>.) The timing of the footfalls is close to 0 and 1/4 of the period of this rhythmic motion. The primitive ricocheting jump of a Norway rat and an Asia Minor gerbil also has this cadence⁶.

Prediction 3. We next predict the occurrence of half-integer wave numbers in myriapod gaits. Because the network has twice as many cells as the animal has legs, the number of waves of leg movement that fit into the observable half of the network is either an integer or half an odd integer. Manton²⁵ provides drawings of several centipede gaits; the number of waves is close either to an integer (4, 3, 2) or half an odd integer (3/2); see Fig. 2. The tripod gait of hexapods²⁶ is a 3/2 wave.

Prediction 4. Finally, we specialize the network to bipeds where hidden cells seem unnecessary⁹. For evolutionary reasons, we expect bipeds not to break the pattern hypothesized for many-legged animals, thus resulting in four primary bipedal gaits (Table 2). If hidden cells do occur in bipedal CPGs, they should play an active role, perhaps controlling the timing of different muscle groups. Thus, muscle groups may reveal the presence of two distinct gaits in which the legs move half a period out of phase. More precisely, lower leg muscles should be activated synchronously in one gait and asynchronously in the other—because of the phases of the hidden cells. The human gaits walk and run support this prediction^{27,28}. During walking, the gastrocnemius (an ankle plantarflexor) and the tibialis anterior (an ankle dorsiflexor) are activated out of phase, whereas during running, they are co-activated during significant portions of the gait cycle.

Unlike bipeds, the hidden phases of quadrupedal primary gaits can be deduced from the observable half network. Thus, each primary gait corresponds either to synchronous muscle activation (as in the run) asynchronous activation (as in the walk). This prediction should be testable. □

Received 21 May; accepted 9 August 1999.

- Delcomyn, F. Neural basis of rhythmic behavior in animals. *Science* **210**, 492–498 (1980).
- Grillner, S. Locomotion in vertebrates: central mechanisms and reflex interaction. *Physiol. Rev.* **55**, 247–304 (1975).
- Selverston, A. I. Are central pattern generators understandable? *Behav. Brain Sci.* **3**, 535–571 (1980).
- Shik, M. L. & Orlovsky, G. N. Neurophysiology of locomotor automatism. *Physiol. Rev.* **56**, 465–501 (1976).
- Alexander, R. McN. in *Mechanics and Energetics of Animal Locomotion* (eds Alexander, R. McN. & Goldspink, J. M.) 168–203 (Chapman and Hall, London, 1977).
- Gambaryan, P. P. *How Mammals Run: Anatomical Adaptations* (Wiley, New York, 1974).
- Hildebrand, M. Symmetrical gaits of horses. *Science* **150**, 701–708 (1965).
- Hildebrand, M. The quadrupedal gaits of vertebrates. *Bioscience* **39** (11) 766–775 (1989).
- Collins, J. J. & Stewart, I. Coupled nonlinear oscillators and the symmetries of animal gaits. *J. Nonlin. Sci.* **3**, 349–392 (1993).

- Golubitsky, M., Stewart, I., Buono, P. L. & Collins, J. J. A modular network for legged locomotion. *Physica D* **115**, 56–72 (1998).
- Schöner, G., Jiang, W. Y. & Kelso, J. A. S. A synergetic theory of quadrupedal gaits and gait transitions. *J. Theor. Biol.* **142**, 359–391 (1990).
- Collins, J. J. & Stewart, I. Symmetry-breaking bifurcation: a possible mechanism for 2:1 frequency-locking in animal locomotion. *J. Math. Biol.* **30**, 827–838 (1992).
- Collins, J. J. & Stewart, I. Hexapodal gaits and coupled nonlinear oscillator models. *Biol. Cybern.* **68**, 287–298 (1993).
- Collins, J. J. & Stewart, I. A group-theoretic approach to rings of coupled biological oscillators. *Biol. Cybern.* **71**, 95–103 (1994).
- Kopell, N. & Ermentrout, G. B. Symmetry and phaselocking in chains of weakly coupled oscillators. *Comm. Pure Appl. Math.* **39**, 623–660 (1986).
- Kopell, N. & Ermentrout, G. B. Coupled oscillators and the design of central pattern generators. *Math. Biosci.* **89**, 14–23 (1988).
- Kopell, N. & Ermentrout, G. B. Phase transitions and other phenomena in chains of oscillators. *SIAM J. Appl. Math.* **50**, 1014–1052 (1990).
- Buono, P.-L. & Golubitsky, M. Models of central pattern generators for quadruped locomotion: I. primary gaits. *J. Math. Biol.* (submitted).
- Golubitsky, M., Stewart, I. N. & Schaeffer, D. G. *Singularities and Groups in Bifurcation Theory* Vol. 2 275–282 387–399 (Springer, New York, 1988).
- Buono, P.-L. *A Model of Central Pattern Generators for Quadruped Locomotion*. PhD Thesis, Mathematics Department, Univ. Houston (1998).
- Hildebrand, M. in *Neural Control of Locomotion* (eds Herman, R. M., Grillner, S., Stein, P. S. G. & Stewart, D. G.) 203–236 (Plenum, New York, 1976).
- Alexander, R. McN. & Jayes, A. S. A dynamic similarity hypothesis for the gaits of quadrupedal mammals. *J. Zool.* **201**, 135–152 (1983).
- Leach, D. H. & Spriggs, E. Gait fatigue in the racing thoroughbred. *J. Equine Med. Surg.* **3**, 436–443 (1979).
- Deuel, N. R. & Lawrence, L. M. Laterality in the gallop gait of horses. *J. Biomech.* **20**, 645–649 (1987).
- Manton, S. M. The evolution of arthropod locomotory mechanics. Part 8: Functional requirements and body design in Chilopoda. *J. Linnean Soc. (Zool.)* **45**, 251–484 (1965).
- Full, R. J., Blickhan, R. & Ting, L. H. Leg design in hexapodal runners. *J. Exp. Biol.* **158**, 369–390 (1991).
- Mann, R. A. in *Disorders of the Foot* (ed. Jahss, M. H.) 37–67 (W. B. Saunders, Philadelphia, 1982).
- Mann, R. A., Moran, G. T. & Dougherty, S. E. Comparative electromyography of the lower extremity in jogging, running and sprinting. *Am. J. Sports Med.* **14**, 501–510 (1986).

Acknowledgements

We thank A. Golubitsky for help in analysing the jump gait from the rodeo video, kindly supplied by the Houston Livestock Show and Rodeo. We also thank Ray Glantz for discussions. This research was supported in part by grants from the NSF and the Texas Advanced Research Program.

Correspondence and requests for materials should be addressed to M.G. (e-mail: mg@uh.edu).

Probing the human stereoscopic system with reverse correlation

Peter Neri, Andrew J. Parker & Colin Blakemore

University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK

Our two eyes obtain slightly different views of the world. The resulting differences in the two retinal images, called binocular disparities, provide us with a stereoscopic sense of depth¹. The primary visual cortex (V1) contains neurons that are selective for the disparity^{2–4} of individual elements in an image, but this information must be further analysed to complete the stereoscopic process^{5,6}. Here we apply the psychophysical technique of reverse correlation⁷ to investigate disparity processing in human vision. Observers viewed binocular random-dot patterns, with 'signal' dots in a specific depth plane plus 'noise' dots with randomly assigned disparities. By examining the correlation between the observers' ability to detect the plane and the particular sample of 'noise' disparities presented on each trial, we revealed detection 'filters', whose disparity selectivity was remarkably similar to that of individual neurons in monkey V1. Moreover, if the noise dots were of opposite contrast in the two eyes, the

tuning inverted, just like the response patterns of V1 neurons^{5,6}. Reverse correlation appears to probe disparity processing at the earliest stages of binocular combination, prior to the generation of a full stereoscopic depth percept.

Subjects viewed random-dot stereograms⁸ that depicted a background surface with the same disparity as the binocular fixation point, and a central square that was either in the same plane as the background (zero disparity) or standing forward from the background (target disparity). Their task was to report the absence or presence of the target plane (Fig. 1a). An additional disparity drawn from a discrete uniform distribution (a random value within a certain fixed range of disparities; Fig. 1b and Methods) was added to a fraction of the dots in the central square region, whether this region stood forward or not. This made the task much more difficult, requiring the observer to make perceptual decisions about the depth of the signal dots in the central region in the presence of a cloud of interfering noise dots, all at different depths.

A preliminary experiment determined that adding such disparity noise to 95% of the dots in the central region brought detection performance down to about 75% correct. A large number of trials (10,000–15,000) were conducted with these near-threshold stimuli. The actual profile of disparity values in the central region varied from trial to trial owing to the random sampling procedure: some of these variations in the noise distribution had no effect on detection performance, but others were highly influential. We hypothesized that configurations of the noise that affected the detection task would reveal the properties of mechanisms normally involved in the processing of disparity.

After each trial, the computer recorded the number of noise dots at each disparity level, the presence or absence of the target and the observer's response. These individual noise distributions were separately averaged for each possible class of response—hit, false alarm, miss and correct rejection (Fig. 1c and Methods; see also ref. 7). Within each noise pattern, 50% of the dots had same-contrast polarity, dark or bright, in both eyes (binocularly correlated) and 50% had opposite-contrast polarity in each eye (binocularly anti-correlated). The four averages of noise distributions were computed separately for binocularly correlated and anti-

correlated noise dots. (We obtained results similar to those presented here when the effect of each type of binocular correlation was explored independently in individual blocks of trials.) The process of computing the averages separately for each class of psychophysical response allowed us to analyse whether the particular distribution of disparity noise on each trial assisted or hindered the detection of the stereo target. Not surprisingly, 'yes' responses (hits and false alarms) were more common, and 'no' responses less frequent, when there was, by chance, an excess of same-contrast noise dots close to the target depth plane.

Figure 2 shows the summed profiles of noise as a function of disparity for both same-contrast (correlated) and reversed-contrast (anti-correlated) noise dots, for two observers. Correlated noise yields a function with a positive peak at the disparity of the target and clear negative flanks at each side, broadly similar to those derived with other psychophysical techniques^{9,10}. Binocular neurons in primate V1 produce similar functions when their firing rate is measured for stimuli of different disparity^{2–4}. Interestingly, the summed profile for anti-correlated noise (reversed-contrast) generates a roughly opposite profile: there is a small, but distinct, trough at the disparity of the target. Many V1 neurons respond in this way to anti-correlated random-dot stereograms: their response profile is inverted, but of smaller modulation, compared with that for patterns with same contrast in each eye⁶. It is interesting that anti-correlated stimuli can drive vergence eye movements in the geometrically inappropriate direction: an anti-correlated target that steps away from the observer in terms of retinal disparity actually induces a transient convergent eye movement¹¹.

Disparity-selective neurons in the striate cortex have been modelled as units that correlate one-dimensional Gabor-filtered images arriving from each eye^{5,6}. We subjected a synthetic detector of this type to the same experimental procedure used in the psychophysical experiment (Fig. 3a), using exactly the same values for the parameters. No rescaling was applied to the resulting simulations of the detection curves. Figure 3b shows how this model captures all aspects of the psychophysical results, with one noticeable exception: the simulated amplitude of the function for anti-correlated noise is larger than experimentally observed. In reality, this discrepancy

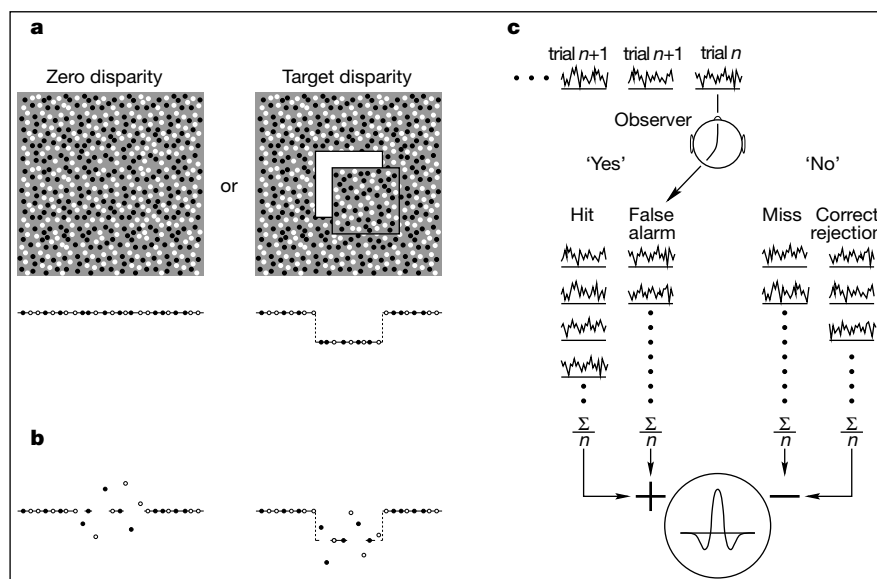


Figure 1 Psychophysical task and reverse-correlation technique. **a**, Random-dot stereograms (front views and top views shown) either contained or did not contain a disparity-defined central square (target). **b**, Observers were required to detect the target, while disparity noise was added to the central region. **c**, Noise configurations for a sequence of threshold stimuli (top row: noise density * as a function of disparity) were

classified according to the observer's response. Summation of average noise distributions yielding 'yes' responses (hits and false alarms) and subtraction of those giving 'no' responses (misses and correct rejections) generates the presumed disparity-selective sensory filter (response versus disparity).

highlights the similarity of psychophysical and neurophysiological results: neurons in monkey V1 and human psychophysical detection functions both exhibit a relatively reduced amplitude of modulation for anti-correlated stimuli. Qualitatively, the inversion of the function for reversed-contrast stimuli is expected from any mechanism that preserves the polarity of the Gabor-filtered inputs. The smaller amplitude might be explained by some threshold or saturating nonlinearity in the contrast-response functions of the underlying monocular filters. For binocularly correlated noise, this simple model gives reasonably accurate predictions of human performance, chiefly because we introduced a large amount of external noise, which dominates the statistical level of performance for both human observers and the model. As the synthetic unit did not have any internal noise, one can conclude that the effect of any internal neural noise in the human observers was also relatively slight in these circumstances.

These results provide important clues about the way in which the neural processing of binocular disparity leads to the perception of stereoscopic depth. Dense anti-correlated random-dot stereograms elicit responses (albeit inverted in modulation) from binocular neurons in early visual areas⁶ but such stimuli do not support stereoscopic vision^{8,12–14}. Sparse stereograms and single bars of opposite contrast in the two eyes can provide a weak impression of depth^{12,13}, but when stereopsis occurs, it is at a depth consistent with the retinal disparity of the contrast-reversed elements¹⁴. From these perceptual observations, one would expect the anti-correlated response function in the present experiments to be qualitatively similar in form to that for correlated noise, rather than inverted as in Fig. 2.

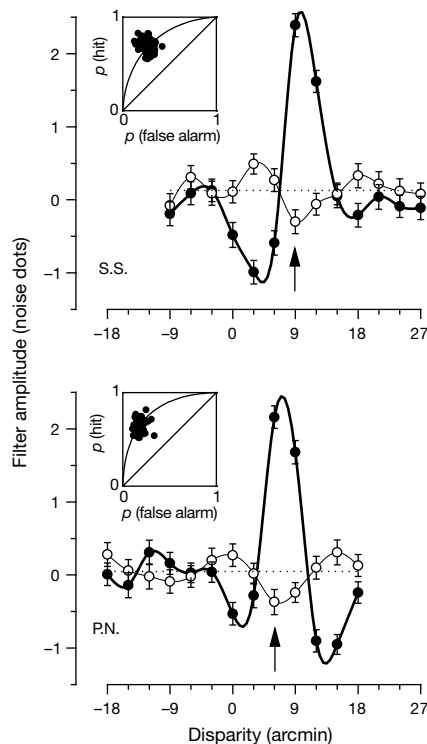


Figure 2 Tuning functions for two observers (top, observer S.S., 11,000 trials; bottom, observer P.N., 13,000 trials). Solid symbols (± 1 s.e.) and thick splines plot response functions for binocular contrast-correlated noise; open symbols (thin splines) plot functions for anti-correlated noise. Dotted lines are mean values. Arrows indicate target disparity: 9 arcmin for S.S. and 6 arcmin for P.N. Correlated noise yields a Mexican-hat-shaped function; anti-correlated noise an inverted pattern of smaller amplitude. The insets plot probability of hit against probability of false alarm; each dot refers to an individual block. Smooth curves plot a constant level of performance in terms of d' (see Methods). Average d' values were 1.1 (S.S.) and 1.3 (P.N.).

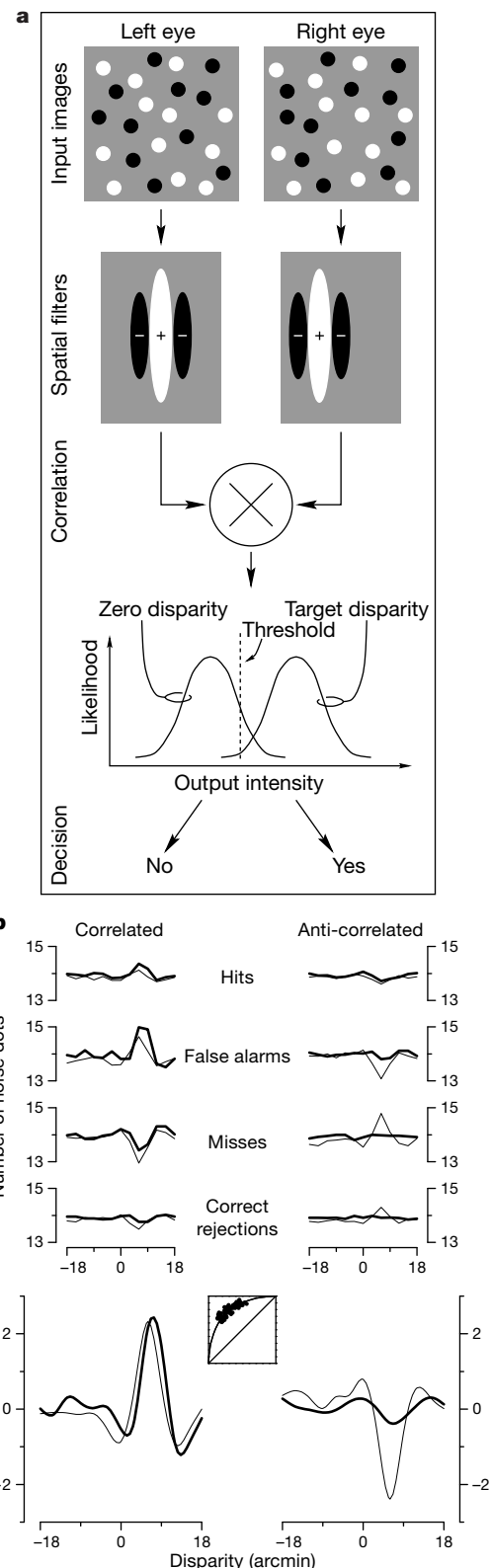


Figure 3 Model and simulation of data. **a**, The synthetic detector consisted of two monocular spatial filters, offset to a characteristic preferred disparity, whose outputs were correlated to yield the response^{5,6}. This was converted to a binary choice. **b**, The detector was challenged with stimulus images used for the psychophysics, and the reverse-correlation technique applied. Thick lines summarize experimental data for subject P. N.; thin lines are the corresponding simulation curves. Average noise configurations for all four stimulus-response classes are shown (upper part), as is the combined disparity-tuning function (lower part). The model reproduces most experimental features (including d' scores, shown in the central inset: conventions as in Fig. 2), apart from the reduced amplitude of the function for anti-correlated noise.

This raises the following question: given that anti-correlated stimuli alone lead to either a weak percept of correct depth or a percept of no depth at all, how is it that anti-correlated noise reduces the detectability of binocularly correlated stimuli when at the same disparity and enhances detectability when at neighbour disparities? The answer may lie in the stage of processing that is probed by the detection task used here. At the earliest stage of binocular combination, inputs from the left and right eyes seem to be combined almost linearly^{5,6}. Such a mechanism would provide for an interaction between the binocularly correlated stimulus and the anti-correlated noise dots. The data presented here show that this stage can be identified psychophysically by the reverse-correlation technique⁷. This powerful new tool may be of general value in revealing the properties of low-level perceptual mechanisms.

Our results provide insights into how the perception of depth arises in the human visual system. We can now identify at least two separate stages in the generation of a full stereoscopic percept. In the initial stage, information from the two eyes is combined in a local, feature-specific manner². This process is close to a measurement of binocular correlation, thus exhibiting an inverted response when anti-correlated stimuli are used. This stage, perhaps occurring in human V1, appears to support the basic detection task in this study. At a later stage, the detailed matching of information between the two eyes is sorted out on a broader scale. The overall binocular correspondence problem is solved to yield the perception of surfaces and boundaries in depth from potentially multiply ambiguous local cues^{15,16}. If our speculation that V1 holds the initial stage of binocular combination is correct, it remains to identify the site or sites where true stereoscopic correspondence is established and signals from dense, globally anti-correlated patterns are rejected. □

Methods

Stimuli were generated by a VSG graphics card (CRS, Rochester, UK) and presented on a CRT monitor (Vision Master 17, Iiyama) at a viewing distance of 57 cm. To provide independent stimulation of the eyes, the graphics card was synchronized with ferroelectric stereo-goggles at 110 Hz (55 Hz for each eye). The observer fixated a central red cross, presented to both eyes (zero disparity), against a grey background (35 cd m⁻²). This cross disappeared during the 250-ms presentations of a stereogram subtending 11.5° × 11.5°. Each stereogram consisted of randomly distributed dots (6 arcmin diameter; density 12%), half of which were bright (65 cd m⁻²) and half dark (5 cd m⁻²), on the same grey background so that stimulus onset did not change mean luminance.

The periphery of all stereograms consisted entirely of correlated dots at zero disparity, perceived as a background surface. Within a central square (5.8° × 5.8°), the signal dots all had the same disparity, either zero (forming a surface continuous with the background), or a target disparity (6 arcmin crossed disparity for one subject, 9 arcmin for the other, corresponding to a square surface, closer than the background). The disparity of the other dots in the central region was drawn randomly from a discrete uniform distribution, spanning 13 disparity levels of 3 arcmin steps, centred around either the target disparity (subject S.S.) or zero disparity (subject P.N.). Half the noise dots in each stereogram were matched in contrast in the two eyes (both either dark or bright); the others were of opposite contrast (one dark and one bright). They were painted alternately, so that there was an equal probability that correlated noise dots would be occluded by anticorrelated and vice versa. Signal dots were evenly interspaced among noise dots in the drawing sequence.

At the end of each presentation, the observer made a forced-choice decision about whether signal dots were at the target disparity. This decision triggered the next presentation after a 500-ms delay. Every 30–50 trials, a target square without noise was presented so as to remind the observer of its appearance and to help to maintain vigilance. Subject S.S. was naive of the goals and methodology of the study.

Estimation of noise tolerance threshold

In signal-detection theory¹⁷, the parameter d' provides a dimensionless measure of the detectability of a stimulus, expressing signal strength as a multiple of the standard deviation of the underlying probability distribution of the detection process. During these experiments, the stimulus always had 95% noise dots and 5% signal dots, which yielded $d' \sim 1.2$. These values were estimated from a preliminary block of ~200 trials during which the fraction of noise dots was adjusted up or down after each trial¹⁸.

Reverse-correlation analysis

Our technique was adapted from that developed by Beard and Ahumada⁷ for a Vernier task. Blocks consisted of 200 trials. The disparity distribution of noise dots for each presentation was stored and sorted according to which response class had occurred (hit, false alarm, miss or correct rejection). Noise distributions were averaged to obtain the

mean noise configuration associated with each response. By summing the average configurations for the 'yes' responses (hits and false alarms) and subtracting those for the 'no' responses (misses and correct rejections), we obtained the single tuning function for the presumed sensory filter used to perform the task (Fig. 1c).

Estimates of variance for the data points in Fig. 2 were obtained with a bootstrap method¹⁹, in which random picks (with replacement) were repeatedly taken from the experimentally obtained data sets according to a pattern that matched the actual distribution of classes of response produced during the experiment. 1,000 samples of these synthetic data were subjected to the same analysis procedure used for the true data. This generates a bootstrap estimate of the sampling distribution of the values plotted in Fig. 2. The error bars show ± 1 s.e.

Modelling

We used a synthetic detector similar to that of Ohzawa *et al.*⁵, but with the two monocular receptive fields extended to both spatial dimensions (Fig. 3a). Stimulus images, generated exactly as for the experiments, were filtered by x -Gabor/ y -gaussian filters, which had a disparity offset (the preferred disparity for the detector), and sampled the image at a number of different positions (i, j in the equation below). We chose a position-disparity model (as opposed to phase disparity) because the spatial frequency of the filter (see below) was well above 2.5 cycles per deg (ref. 20).

$$\text{Output} = \sum_{i,j} \left[\sum_{x,y} f_R(x,y) I_R(x+i, y+j) \right] \cdot \left[\sum_{x,y} f_L(x,y) I_L(x+i, y+j) \right]$$

where $I(i, j)$ is the image, R and L refer to the eyes, and the filter f is of the form:

$$f(x, y) = \cos(\omega x) \cdot G(x, \sigma_x) \cdot G(y, \sigma_y)$$

where G is a gaussian function of standard deviation σ ($\sigma_x = 3$ arcmin, $\sigma_y = 1.5^\circ$, $\omega = 5$ cycles per deg for Fig. 3b). The output was converted into a yes–no response by a statistical thresholding procedure (Fig. 3a, bottom). The reverse-correlation technique was applied to the outputs of this synthetic detector to generate response functions like those obtained in the psychophysical experiments.

Received 7 May; accepted 16 August 1999.

- Wheatstone, C. Contributions to the physiology of vision—Part the first. On some remarkable and hitherto unobserved phenomena of binocular vision. *Phil. Trans. R. Soc. Lond.* **128**, 371–394 (1838).
- Barlow, H. B., Blakemore, C. & Pettigrew, J. D. The neural mechanism of binocular depth discrimination. *J. Physiol. (Lond.)* **193**, 327–342 (1967).
- Nikara, T., Bishop, P. O. & Pettigrew, J. D. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp. Brain Res.* **6**, 353–372 (1968).
- Poggio, G. F. & Fischer, B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving Rhesus monkey. *J. Neurophysiol.* **40**, 1392–1407 (1977).
- Ohzawa, I., DeAngelis, G. C. & Freeman, R. D. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* **249**, 1037–1041 (1990).
- Cumming, B. G. & Parker, A. J. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* **389**, 280–283 (1997).
- Beard, B. L. & Ahumada, A. J. Jr in *Human Vision and Electronic Imaging III* (eds Rogowitz, B. E. & Pappas, T. N.) *Proc. SPIE* **3299**, 79–85 (1998).
- Julesz, B. *Foundations of Cyclopean Perception* (Univ. Chicago Press, 1971).
- Stevenson, S. B., Cormack, L. K., Schor, C. M. & Tyler, C. W. Disparity tuning in mechanisms of human stereopsis. *Vision Res.* **32**, 1685–1694 (1992).
- Cormack, L. K., Stevenson, S. B. & Schor, C. M. Disparity-tuned channels of the human visual system. *Vis. Neurosci.* **10**, 585–596 (1993).
- Masson, G. S., Busettini, C. & Miles, F. A. Vergence eye movements in response to binocular disparity without depth perception. *Nature* **389**, 283–286 (1997).
- Cogan, A. I., Kontsevich, L. L., Lomakin, A. J., Halpern, D. L. & Blake, R. Binocular disparity processing with opposite-contrast stimuli. *Perception* **24**, 33–47 (1995).
- Cogan, A. I., Lomakin, A. J. & Rossi, A. F. Depth in anti-correlated stereograms: effects of spatial density and interocular delay. *Vision Res.* **33**, 1959–1975 (1993).
- Cumming, B. G., Shapiro, S. E. & Parker, A. J. Disparity detection in anticorrelated stereograms. *Perception* **27**, 1367–1377 (1998).
- Marr, D. & Poggio, T. A computational theory of human stereo vision. *Proc. R. Soc. Lond. B* **204**, 301–328 (1979).
- Fleet, D. J., Wagner, H. & Heeger, D. J. Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res.* **36**, 1839–1857 (1996).
- Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
- Watson, A. B. & Pelli, D. G. QUEST: a Bayesian adaptive psychometric method. *Percept. Psychophys.* **33**, 113–120 (1983).
- Efron, B. & Tibshirani, R. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
- Schor, C., Wood, I. & Ogawa, J. Binocular sensory fusion is limited by spatial resolution. *Vision Res.* **24**, 661–665 (1984).

Acknowledgements

We thank H. Barlow and B. Cumming for comments. This work was supported by the Oxford McDonnell–Pew Centre for Cognitive Neuroscience, the Medical Research Council and the Wellcome Trust.

Correspondence and requests for materials should be addressed to P.N. (e-mail: peter.neri@physiol.ox.ac.uk).

Top-down signal from prefrontal cortex in executive control of memory retrieval

Hyoe Tomita*, Machiko Ohbayashi*, Kiyoshi Nakahara†, Isao Hasegawa*‡ & Yasushi Miyashita*†‡

* Department of Physiology, The University of Tokyo, School of Medicine, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

† Mind Articulation Project, ICORP, Japan Science and Technology Corporation, Bunkyo-ku, Tokyo 113-0034, Japan

‡ National Institute for Physiological Sciences, Myodaiji-cho, Okazaki 444-8585, Japan

Knowledge or experience is voluntarily recalled from memory by reactivation of the neural representations in the cerebral association cortex^{1–4}. In inferior temporal cortex, which serves as the storehouse of visual long-term memory^{5–8}, activation of mnemonic engrams through electric stimulation results in imagery recall in humans⁹, and neurons can be dynamically activated by the necessity for memory recall in monkeys^{10,11}. Neuropsychological studies¹² and previous split-brain experiments¹³ predicted that prefrontal cortex exerts executive control upon inferior temporal cortex in memory retrieval; however, no neuronal correlate of this process has ever been detected. Here we show evidence of the top-down signal from prefrontal cortex. In the absence of bottom-up visual inputs, single inferior temporal neurons were activated by the top-down signal, which conveyed information on semantic categorization imposed by visual stimulus–stimulus association. Behavioural performance was severely impaired with loss of the top-down signal. Control experiments confirmed that the signal was transmitted not through a subcortical but through a fronto-temporal cortical pathway. Thus, feedback projections from pre-

frontal cortex to the posterior association cortex^{2,3,14} appear to serve the executive control of voluntary recall.

We conducted single-unit recording experiments using the posterior-split-brain paradigm^{13,15} which was originally introduced in humans^{12,16} (Fig. 1). Two monkeys (*Macaca fuscata*) underwent transection of the posterior corpus callosum and anterior commissure, leaving the anterior corpus callosum which interconnects the prefrontal cortices^{13,15} (Fig. 1c–e). In this preparation, inferior temporal neurons in one hemisphere ('electrode', Fig. 1a) are activated by bottom-up visual inputs when an object is presented in the visual hemifield contralateral to the recording site^{17,18}. When the object is presented in the ipsilateral hemifield, however, these neurons do not receive bottom-up visual inputs; any neural activation should reflect top-down inputs from the prefrontal cortex¹³ (Fig. 1b).

The monkeys were trained to memorize visual stimulus–stimulus associations among 20 cue and 5 choice pictures^{10,13}. For each trial, while the monkey held a lever and maintained fixation, cue and choice pictures were sequentially presented parafoveally with a delay of 600 to 1,500 ms (Fig. 1a, b). If the choice picture was the correct one associated with the cue, the monkey had to release the lever. A trial was aborted if the monkey's gaze was more than 0.6° away from the fixation point, which ensured that the visual sensory input was confined to one hemisphere.

Figure 2a shows responses of an inferior temporal neuron, which was not only activated by contralateral presentation of stimuli (bottom-up response, black) but was also activated by ipsilateral presentation of stimuli (top-down response, blue). The cue that elicited the optimal bottom-up response (thick black) also elicited the strongest top-down response when presented in the ipsilateral hemifield (thick blue). The cue that did not elicit the bottom-up response (thin black) did not elicit a top-down response either (thin blue). The stimulus selectivity of the top-down response was correlated with that of the bottom-up response (Fig. 2b; Pearson's correlation coefficient, $r = 0.788$, $P < 0.001$). The latency of the top-down response (178 ms) was longer than that of the bottom-up

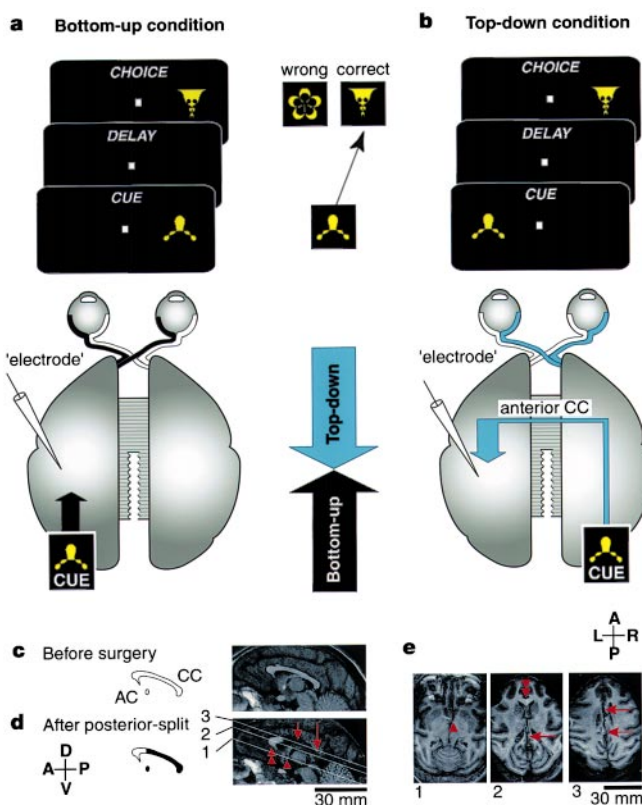


Figure 1 Experimental design. **a**, Bottom-up condition. Visual stimuli (cue and choice pictures) were presented in the hemifield contralateral to the recording site ('electrode') in the inferior temporal cortex. The monkey must choose the correct choice specified by the cue. Fixation was required throughout a trial ($<0.6^\circ$). Bottom-up sensory signals (black arrow) would be detected in this condition. **b**, Top-down condition. As in **a**, but the cue was presented in the hemifield ipsilateral to the recording site, whereas the choice was presented contralaterally. In posterior-split-brain monkeys (see **d**, **e**), sensory signal does not reach visual areas in the opposite hemisphere. In this condition, only top-down signals (blue arrow) could activate inferior temporal neurons through feedback connections from the prefrontal cortex. **c**, **d**, Mid-sagittal MRI slices of the monkey brain before (**c**) and after (**d**) transection of posterior corpus callosum (CC, arrows) and anterior commissure (AC, arrowhead). Remaining part of CC is marked by a double arrowhead. Extent of lesion is schematically drawn (filled) to the left. **e**, Horizontal slices at the levels indicated in **d**. Posterior CC (arrows) and AC (arrowhead) were transected. Anterior CC (double arrowhead) was intact. Horizontal MR images were slightly distorted at the sites of head implants.

response (73 ms). We recorded from 543 neurons in the inferior temporal cortex of the two monkeys at this 'posterior-split' stage, and found 435 neurons to show task-related activity. We further assessed the stimulus selectivity by performing analysis of variance (ANOVA) in 104 task-related neurons which were tested in more than 3 trials for each of 20 cues in both the top-down and bottom-up conditions. Among them, 73 and 45 cells showed significantly stimulus-selective activity in the bottom-up and the top-down conditions, respectively. Forty-three cells were significantly stimulus-selective in both of these conditions. The ensemble average of activities of these 43 neurons showed robust top-down responses to the optimal cue picture in the bottom-up condition (Fig. 2c). The distribution of the correlation coefficients between top-down and bottom-up responses for each of the 43 neurons (median 0.54) was positively shifted from the distribution of the baseline correlations (median -0.00) estimated from Monte Carlo simulations (Fig. 2d, Kolmogorov-Smirnov test, $P < 0.001$), which confirmed similarity of the stimulus selectivity in the top-down and bottom-up conditions. In 36 of the 43 cells, we could determine unambiguous onset of both top-down and bottom-up responses and found that the latency was significantly longer in the top-down condition (Fig. 2e; two-tailed t -test, $P < 0.001$).

The top-down signal probably reaches the inferior temporal cortex through the prefrontal cortex, which sends rich backwards

projections to the inferior temporal cortex¹⁹. To exclude the possibility that the activation observed here reflected indirect inputs from the subcortical structures²⁰, the neural response was examined after further transection of the remaining anterior corpus callosum ('full-split' stage, Fig. 3a). At the full-split stage, 97 neurons were recorded and 88 were task-related. Among them, 43 neurons were tested for their stimulus selectivities both in the top-down and the bottom-up conditions. While 28 of them showed significant stimulus-selectivity in the bottom-up condition, none did so in the top-down condition, which was in significant contrast to the posterior-split stage ($P < 10^{-10}$, Fig. 3b). Averaged activity of the 28 neurons showed that the top-down response was abolished after the full-split surgery (Fig. 3c). These observations confirmed that the top-down activation at the posterior-split stage reflected the signal from the prefrontal cortex. After the full-split surgery, behavioural performance was severely impaired in the top-down condition (Fig. 3d), which implied behavioural relevance of the top-down signals.

Our category-association task (Fig. 4a) allows further characterization of the top-down signals. We found category-selective responses during the delay interval (Fig. 4b): delay activity was raised for all cues in Category I, but not for any cues in Category V. ANOVA confirmed that the delay activity was significantly categorical ($F[4, 15] = 6.21$, $P < 0.01$; with *post-hoc* Tukey's test,

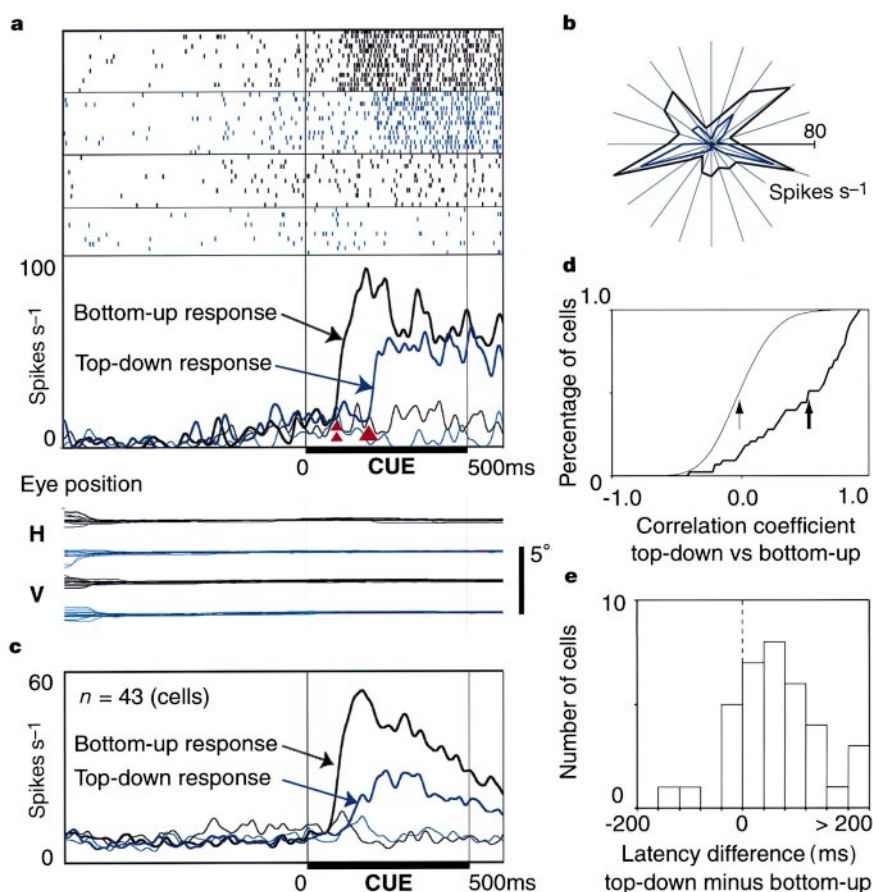


Figure 2 Neuronal activity in top-down condition. **a**, Single inferior temporal cell (top-down, blue; bottom-up, black). Raster displays, spike density functions (SDFs) and eye position traces were aligned at the cue onset. In the SDFs, thick lines show responses to the optimal cue, whereas thin lines show responses to a null cue. Onset of the top-down response (arrowhead) was later than that of the bottom-up response (double arrowhead). Horizontal (H) and vertical (V) eye positions in all the trials for the optimal cue are shown. **b**, Stimulus selectivity of the cell shown in **a**. Responses to 20 cues were highly selective both for bottom-up condition (ANOVA, $F[19, 105] = 35.7$, $P < 0.0001$, black) and top-

down condition ($F[19, 108] = 17.4$, $P < 0.0001$, blue). **c**, Averaged responses of 43 neurons activated by top-down signals. SDFs are denoted as in **a**. Top-down responses were collected for the cues which elicited best/worst responses in the bottom-up condition. **d**, Cumulative histogram of response correlations between top-down and bottom-up responses for 20 cues. Thick line, experimental data from 43 neurons. Thin line, data from Monte Carlo simulation. Arrows denote median correlation coefficients for experimental data (thick) and simulated data (thin). **e**, Distribution of latency differences (top-down minus bottom-up).

$P < 0.01$). Thirty-four of the 104 cells exhibited selective delay activity in the top-down condition, and 23% of them (8 out of 34) were categorical in the delay interval. The neuron depicted in Fig. 4b also exhibited selective choice response ($F[4, 128] = 7.88$, $P < 0.001$). The strength of the choice response was predictable²¹ from the category-selective delay activity (Fig. 4c), as seen by a significant correlation between them ($r = 0.93$, $P < 0.01$). We also examined the temporal dynamics of the response correlation. The correlation with choice response was developing during the delay interval²¹, and the correlation with cue response was decaying (Fig. 4d, black triangle denotes the 5% significance level, $r = 0.88$). In the 34 neurons with selective top-down delay activity, the distribution of correlation coefficients between delay and choice responses (median 0.41) was significantly different from that of the baseline correlations (median +0.00) obtained by Monte Carlo simulations (Fig. 4e; Kolmogorov-Smirnov test, $P < 0.001$). Thus, the top-down signal triggered development of prospective information encoding the choice picture to be recalled; a recent report on prospective coding in prefrontal cortex also supports this conclusion²¹.

In the present experiment, inferior temporal neurons were activated by top-down signals without bottom-up sensory inputs when monkeys were performing the visual stimulus-stimulus association task (but see ref. 17 for anesthetized monkeys). The longer onset latency of the top-down response (Fig. 2e) would be at least partially ascribed to multisynaptic conduction delay reflecting the signal transformation within the prefrontal cortex²², although it could also be due to the longer accumulation of weaker inputs needed to reach threshold.

In humans, functional neuroimaging studies have revealed that prefrontal areas are activated in various memory retrieval tasks in the absence of external stimuli^{23,24}. In monkeys, disruption of interactions between the prefrontal and inferior temporal cortices through the uncinate fascicle impairs performance of visual stimulus-stimulus association tasks, even for the stimulus sets the monkey had learned pre-operatively^{25,26}. Taken with these observations, our results imply that the prefrontal signal to the temporal cortex contributes to this kind of memory retrieval. Behavioural

relevance of the signal²⁷ can be explored by analyses on error trials, but the number of error trials was not sufficient for such analyses in the present experiment. It will be important to explore whether the prefrontal signal might be a kind of bottom-up signal through another route and might serve the modulation of the visual input to the inferior temporal cortex, or whether it might directly serve memory retrieval by dynamically creating an internal representation of the external world in the posterior association cortex. □

Methods

Animal preparation

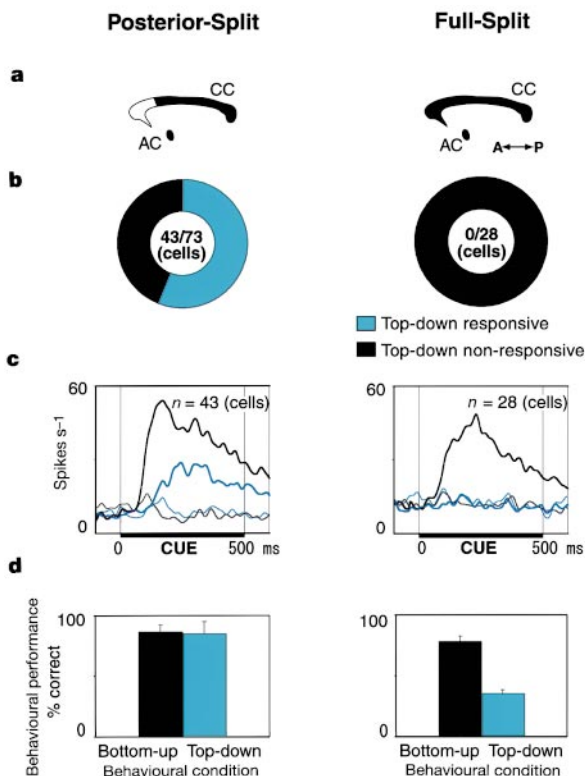
Experiments were conducted in accordance with the NIH Guide for the Care and Use of Laboratory Animals and the regulations of the University of Tokyo School of Medicine. We used two male monkeys (*Macaca fuscata*). In the posterior-split surgery, the posterior corpus callosum was aspirated from the caudal end to the level of the interventricular foramen, where the lateral ventricle was entered and the anterior commissure was cauterized^{13,15} (Fig. 1c–e). The extent of callosal lesions was confirmed by magnetic resonance imaging (MRI) (1.5 T, IR sequence, voxel size = $0.4 \times 0.4 \times 3$ mm³, TR = 2 s, TE = 30 ms, TI = 500 ms). In the full-split surgery, the callosal lesion was extended anteriorly to the rostrum of the corpus callosum. At the end of the experiments, the monkeys were perfused with 4% paraformaldehyde in phosphate buffer (pH 7.4). Adjacent brain sections (50 μ m) were stained with cresyl violet or stained for myelin using the modified Gallyas silver technique¹³, which confirmed that the forebrain commissure was totally split.

Task procedure

Monkeys were trained in a modified visual stimulus-stimulus association task^{10,13} (Fig. 1a, b). Twenty Fourier descriptors extending less than $2^\circ \times 2^\circ$ were used as cues and randomly sorted into five categories. Four cue pictures in each category shared no geometrical similarity and were associated with one common choice picture (Fig. 4a). When the monkey pulled a lever, a trial started and a fixation spot appeared. During fixation ($< 0.6^\circ$), visual stimuli were presented with their centre 2.5° lateral and with their nearest edge more than 1.5° lateral to the fixation spot. The monkey was cued by the first picture (for 500 ms), and after a delay interval (600–1,500 ms), 1–3 choices (500 ms) were sequentially presented. To obtain liquid reward, the monkey was required to release the lever immediately (< 600 ms) after offset of the correct choice. Eye positions were monitored using the scleral search coil method¹³.

Electrophysiology

Extracellular discharges of single neurons were recorded as described^{10,11}. Spike trains were smoothed by convolution with a gaussian kernel ($\sigma = 10$ ms) to obtain spike density



◀ **Figure 3** Comparison before (left) and after (right) the full-split surgery. **a**, Schematic drawings of commissurotomy. Filled areas denote extent of the lesion. **b**, Number of cells responsive in top-down condition (blue) out of those responsive in bottom-up condition. **c**, Averaged responses of the 43 neurons activated by top-down signals. SDFs are denoted as in Fig. 2 (left panel is reproduced from Fig. 2c). Top-down response was abolished after the full-split surgery. **d**, Behavioural performance of the visual association task averaged for all recording sessions. After full-split surgery, the performance was severely impaired in top-down condition (blue), but not in bottom-up condition (black). Error bar, s.d.

function (SDF). The baseline activity was defined as mean discharge rate during the 300-ms period just preceding cue onset. The latency of the bottom-up or top-down response was determined separately as the time point when SDF for the optimal stimulus first exceeded $+2$ s.d. level of baseline activity²⁸. The time window for spike counts was the 300-ms period from the onset of the neural response¹⁴. In case we could not obtain the latency properly we used the default 300-ms time window starting 100 ms after cue onset. Net cue responses were calculated from spike counts by trialwise subtraction of baseline activity to reduce fluctuations that were time correlated for a short period within a trial²⁹. We also

calculated without subtraction and found the statistical results unaltered. Delay activity was collected from the whole delay period without subtraction.

Data analysis

A neuron was classified as task-related if the difference in neural activities in four task periods (fixation, cue presentation, delay interval, choice presentation) was significant (ANOVA, $P < 0.01$). Stimulus selectivity during cue presentation or delay interval was

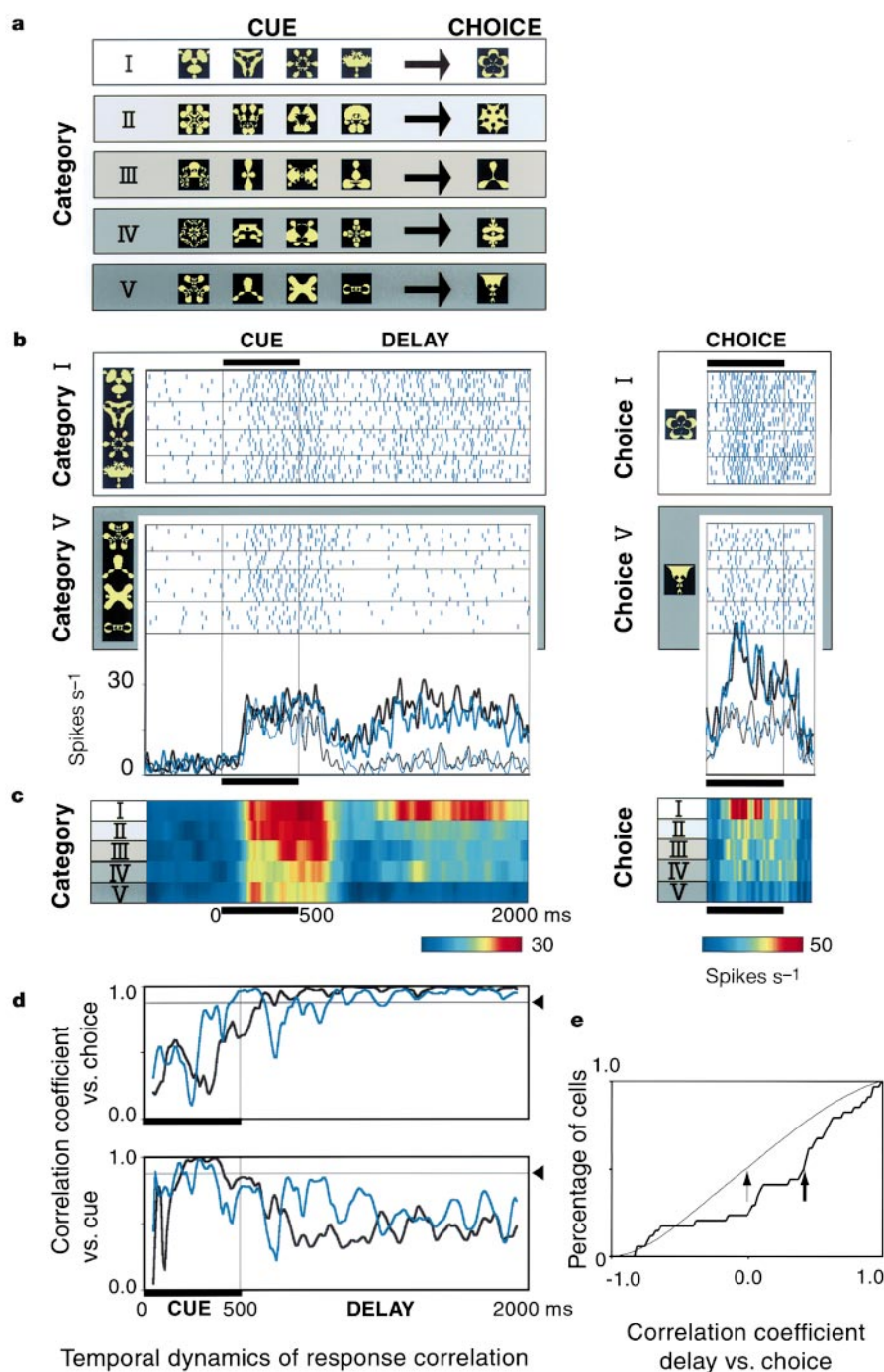


Figure 4 Delay activity of inferior temporal neurons in top-down condition. **a**, Five categories imposed by stimulus–stimulus association. Twenty cue-pictures were randomly sorted into five categories. Each of the four cues in one category specified a common choice. **b**, Category-selective delay activity of an inferior temporal neuron. Delay activities were raised for all cues in Category I, but not for any cues in Category V (rastergrams, top-down condition). Choice responses were also strongest for Category I and weakest for Category V. SDFs show averaged activities across four cues in Category I

(thick) and in Category V (thin) for both conditions (top-down, blue; bottom-up, black). Only the correct trials are shown. **c**, SDFs for five categories shown by a pseudocolor coding. **d**, Temporal dynamics of response correlation. Correlation coefficients between instantaneous firing rates of five categories and corresponding choice (upper) or cue responses (lower) are plotted against the time axis. Triangles indicate the 5% significance level ($r = 0.88$). **e**, Cumulative histogram of response correlations between delay and choice responses ($n = 34$ cells). Symbols are denoted as in Fig. 2d.

tested by ANOVA ($P < 0.05$)^{10,29}. A correlation coefficient between top-down and bottom-up responses was calculated for each single neuron. We performed Monte Carlo simulation by randomizing the assignment of cue stimulus and by recalculating the correlation for 10,000 times in each neuron to estimate a baseline correlation level. The median values of 10,000 simulated correlation coefficients ranged from -0.007 to +0.005 ($n = 43$ cells). Category selectivity was appraised using ANOVA with mean cue responses in each category used as units of analysis and *post-hoc* multiple comparisons (Tukey's method), evaluated at $P = 0.05$. This procedure correctly rejects the case in which neurons are strongly activated only by a single stimulus. Error trials were omitted for the analysis of the choice responses. Trial based instantaneous firing rate (IFR) was defined as the mean discharge rate during the 100-ms window centred at the given time point stepped by 10 ms. Averaging the IFRs first across trials for each cue, then across cues in each category resulted in mean IFRs of five categories.

Received 30 June; accepted 10 August 1999.

- Goldman-Rakic, P. S. in *Handbook of Physiology* Vol. 5 (ed. Plum, F.) 373–417 (American Physiological Society, Bethesda, 1987).
- Fuster, J. M. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe* 3rd edn (Lippincott-Raven, Philadelphia, 1997).
- Petrides, M. in *Handbook of Neuropsychology* (eds Boller, F. & Grafman, J.) 9, 59–82 (Elsevier, Amsterdam, 1994).
- Miyashita, Y. in *The Cognitive Neurosciences* 2nd edn (ed. Gazzaniga, M. S.) in the press (MIT Press, Cambridge, MA).
- Mishkin, M. A memory system in the monkey. *Phil. Trans. R. Soc. Lond. B* **298**, 83–95 (1982).
- Rolls, E. T. Neural organization of higher visual functions. *Curr. Opin. Neurobiol.* **1**, 274–278 (1991).
- Miyashita, Y. Inferior temporal cortex: where visual perception meets memory. *Annu. Rev. Neurosci.* **16**, 245–263 (1993).
- Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
- Penfield, W. & Perot, P. The brain's record of auditory and visual experience. *Brain* **86**, 595–696 (1963).
- Sakai, K. & Miyashita, Y. Neural organization for the long-term memory of paired associates. *Nature* **354**, 152–155 (1991).
- Naya, Y., Sakai, K. & Miyashita, Y. Activity of primate inferotemporal neurons related to a sought target in pair-association task. *Proc. Natl Acad. Sci. USA* **93**, 2664–2669 (1996).
- Gazzaniga, M. S. Principles of human brain organization derived from split-brain studies. *Neuron* **14**, 217–228 (1995).
- Hasegawa, I., Fukushima, T., Ihara, T. & Miyashita, Y. Callosal window between prefrontal cortices: cognitive interaction to retrieve long-term memory. *Science* **281**, 814–818 (1998).
- Chafee, M. V. & Goldman-Rakic, P. S. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* **79**, 2919–2940 (1998).
- Eacott, M. J. & Gaffan, D. Interhemispheric transfer of visual learning in monkeys with intact optic chiasm. *Exp. Brain Res.* **74**, 348–352 (1989).
- Sidtis, J. J., Volpe, B. T., Holtzman, J. D., Wilson, D. H. & Gazzaniga, M. S. Cognitive interaction after staged callosal section: evidence for transfer of semantic activation. *Science* **212**, 344–346 (1981).
- Gross, C. G., Bender, D. B. & Mishkin, M. Contributions of the corpus callosum and the anterior commissure to visual activation of inferior temporal neurons. *J. Neurophysiol.* **131**, 227–239 (1977).
- Doty, R. W., Ringo, J. L. & Levine, J. D. Forebrain commissures and visual memory: a new approach. *Behav. Brain Res.* **29**, 267–280 (1988).
- Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
- Ringo, J. L. & O'Neill, S. G. Indirect inputs to ventral temporal cortex of monkey: the influence of unit activity of alerting auditory input, interhemispheric subcortical visual input, reward, and the behavioral response. *J. Neurophysiol.* **70**, 2215–2225 (1993).
- Rainer, G., Rao, S. C. & Miller, E. K. Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.* **19**, 5493–5505 (1999).
- Goldman-Rakic, P. S. Regional and cellular fractionation of working memory. *Proc. Natl Acad. Sci. USA* **93**, 13473–13480 (1996).
- Buckner, R. L., Raichle, M. E., Miezin, F. M. & Petersen, S. E. Functional anatomic studies of memory retrieval for auditory words and visual pictures. *J. Neurosci.* **16**, 6219–6235 (1996).
- Fletcher, P. C., Shallice, T., Frith, C. D., Frackowiak, R. S. & Dolan, R. J. The functional roles of prefrontal cortex in episodic memory. II. Retrieval. *Brain* **121**, 1249–1256 (1998).
- Eacott, M. J. & Gaffan, D. Inferotemporal-frontal disconnection: the uncinate fascicle and visual associative learning in monkeys. *Eur. J. Neurosci.* **4**, 1320–1332 (1992).
- Gutnikov, S. A., Ma, Y. & Gaffan, D. Temporo-frontal disconnection impairs visual-visual paired association learning but not configural learning in macaca monkeys. *Eur. J. Neurosci.* **9**, 1524–1529 (1997).
- Fuster, J. M., Bauer, R. H. & Jervey, J. P. Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Res.* **330**, 299–307 (1985).
- MacPherson, J. A. & Aldridge, J. W. A quantitative method of computer analysis of spike train data collected from behaving animals. *Brain Res.* **175**, 183–187 (1979).
- Sary, G., Vogels, R. & Orban, G. A. Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* **260**, 995–997 (1993).

Acknowledgements

We thank Y. Naya for technical advice. This work was supported by a grant-in-aid for Specially Promoted Research from the Ministry for Education, Science and Culture of Japan, a grant from the Magnetic Health Science Foundation (Y.M.) and a grant from the Ministry for Education, Science and Culture of Japan (I.H.).

Correspondence and requests for materials should be addressed to H.T.

(e-mail: hyoe@m.u-tokyo.ac.jp) or Y.M. (yasushi_miyashita@m.u-tokyo.ac.jp).

L-type calcium channels and GSK-3 regulate the activity of NF-ATc4 in hippocampal neurons

Isabella A. Graef*†, Paul G. Mermelstein†‡, Kryn Stankunas*, Joel R. Neilson*, Karl Deisseroth‡, Richard W. Tsien‡ & Gerald R. Crabtree*

*Department of Pathology and Department of Developmental Biology, Howard Hughes Medical Institute, and ‡Department of Molecular and Cellular Physiology, Beckman Center for Molecular and Genetic Medicine, Stanford University Medical School, 300 Pasteur Drive, Stanford, California 94305-5426, USA

†These authors contributed equally to this work

The molecular basis of learning and memory has been the object of several recent advances, which have focused attention on calcium-regulated pathways controlling transcription. One of the molecules implicated by pharmacological, biochemical and genetic approaches is the calcium/calmodulin-regulated phosphatase, calcineurin^{1–5}. In lymphocytes, calcineurin responds to specific calcium signals and regulates expression of several immediate early genes by controlling the nuclear import of the NF-ATc family of transcription factors^{6–9}. Here we show that NF-ATc4/NF-AT3 (ref. 10) in hippocampal neurons can rapidly translocate from cytoplasm to nucleus and activate NF-AT-dependent transcription in response to electrical activity or potassium depolarization. The calcineurin-mediated translocation is critically dependent on calcium entry through L-type voltage-gated calcium channels. GSK-3 can phosphorylate NF-ATc4, promoting its export from the nucleus and antagonizing NF-ATc4-dependent transcription. Furthermore, we show that induction of the inositol 1,4,5-trisphosphate receptor type 1 is controlled by the calcium/calcineurin/NF-ATc pathway. This provides a new perspective on the function of calcineurin in the central nervous system and indicates that NF-AT-mediated gene expression may be involved in the induction of hippocampal synaptic plasticity and memory formation.

NF-ATc4 messenger RNA was observed in both murine and rat hippocampus using *in situ* hybridization (Fig. 1a) and polymerase chain reaction with reverse transcriptase (RT-PCR) (Fig. 1b). We tested whether endogenous NF-ATc proteins in neurons would activate NF-AT-dependent transcription by transfecting cultured hippocampal neurons with a reporter plasmid in which NF-AT-binding sites control expression of green fluorescent protein (GFP). This strategy allowed us to test the activity of the endogenous NF-AT transcription complex, which in other cell types is controlled through calcineurin and ras signalling^{6,11}. Cotransfection of constitutively active calcineurin and V12 Ras strongly promoted the expression of the reporter in pyramidal neurons (Fig. 1c). NF-AT-dependent transcription was quantified with a luciferase reporter construct⁸ (Fig. 1d). Three-hour stimulation with the calcium ionophore ionomycin and phorbol-12-myristate-13-acetate (PMA) activated luciferase expression driven by endogenous NF-AT. This activity was blocked by a combination of the calcineurin inhibitors FK506 and cyclosporin A (CsA) (Fig. 1d). The high level of calcineurin expression in hippocampal pyramidal neurons led us to use both drugs. The concentrations of FK506 and CsA used in these studies did not inhibit the expression of a constitutively active luciferase reporter gene (data not shown). The NF-AT reporter was dependent upon endogenous neuronal NF-AT, because deletion of the trimerized NF-AT-binding sites abolished activity (Fig. 1d).

Spontaneous electrical activity was a powerful activator of NF-AT-dependent transcription in hippocampal neurons that had

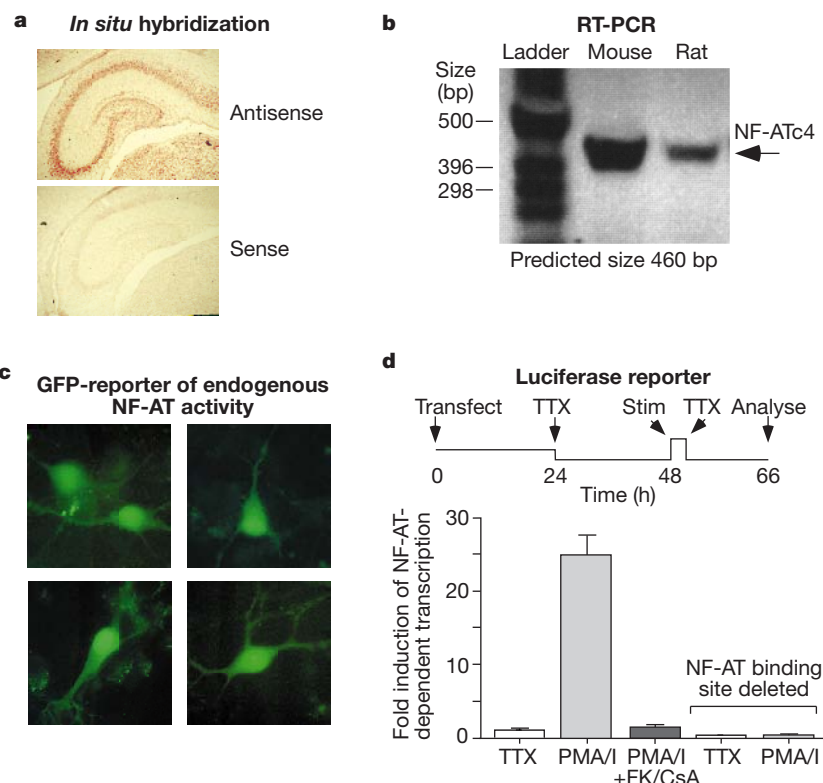


Figure 1 Expression of NF-ATc4 and activation of NF-AT-dependent transcription by endogenous NF-AT proteins in the hippocampus. **a**, *In situ* hybridization of NF-ATc4 messenger RNA in rat hippocampus. **b**, RT-PCR showing NF-ATc4 mRNA in the murine and rat hippocampus. **c**, Cotransfection of constitutively active calcineurin (0.5 μ g) and V12-ras (2 μ g) activated endogenous NF-AT proteins and thereby the expression of an NF-AT-dependent reporter construct (4 μ g) driving the expression of GFP in pyramidal

neurons. **d**, Activation of endogenous NF-AT protein leads to expression of NF-AT-dependent luciferase (4 μ g) after pharmacological stimulation with phorbol-12-myristate-13-acetate (PMA, 25 ng ml⁻¹) and ionomycin (I, 1 μ M) and was blocked by addition of FK506 (200 ng ml⁻¹) and CsA (1 μ g ml⁻¹). A mutant NF-AT-luciferase reporter construct with the trimerized NF-AT binding sites deleted from the minimal IL-2 promoter is no longer activated by PMA and I stimulation.

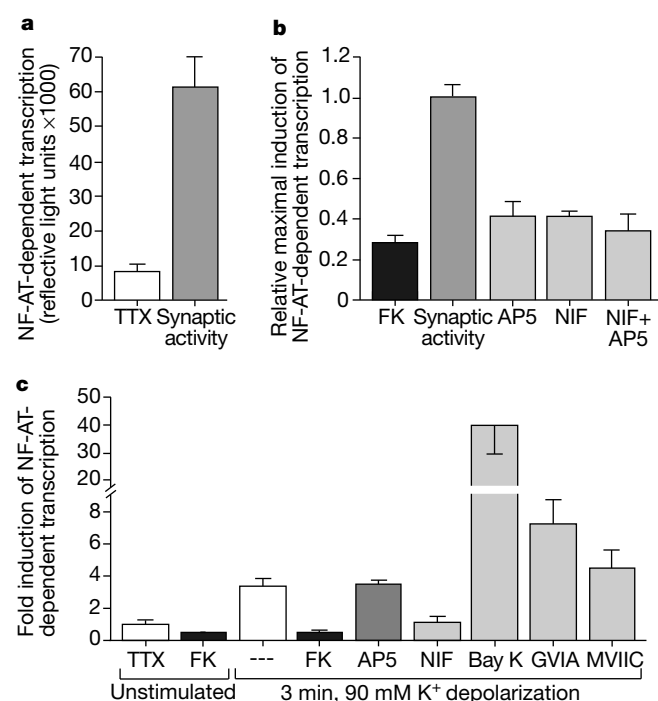


Figure 2 Activation of NF-AT-dependent transcription by NMDA receptors and L-type Ca²⁺ channels. **a**, TTX blocked the activation of NF-AT-dependent transcription by endogenous NF-AT proteins in hippocampal neurons transfected with an NF-AT-luciferase reporter plasmid (4 μ g) at 9 days *in vitro*, a time when synaptic connections had developed³⁰. **b**, FK506/CsA, the L-type channel blocker nifedipine (5 μ M) and AP-5 (25 μ M) inhibit NF-AT-dependent luciferase activity induced by spontaneous electrical activity. **c**, Three minute, 90 mM K⁺ depolarization activated NF-AT-dependent luciferase expression. Transcription was blocked by FK506 and nifedipine, but not by AP-5. (–) Bay K8644 (5 μ M) strongly augmented the transcriptional response. NF-AT-luciferase activity was not blocked by the ω -CgTx-GVIA (1 μ M) or ω -CTx-MVIIC (5 μ M). For the experiments in Fig. 2c, NF-AT-luciferase (4 μ g) and EGFP-NF-ATc4 (1 μ g) were cotransfected.

developed synaptic connections (Fig. 2a). Inhibition of action potentials with the Na⁺-channel blocker tetrodotoxin (TTX) (Fig. 2a) or inhibition of calcineurin by FK506/CsA (Fig. 2b) markedly reduced NF-AT-directed gene expression. To determine whether synaptic activity was required, we applied pharmacological blockers of key postsynaptic neurotransmitter receptors and voltage-gated Ca²⁺ channels (Fig. 2b). The NMDA (N-methyl-D-aspartate)-receptor antagonist AP-5 (25 μ M) and the L-type Ca²⁺ channel blocker nifedipine (5 μ M) each reduced NF-AT-dependent gene expression to an extent not significantly different from the blockers in combination (Fig. 2b). Evidently, activity of both kinds of channels was critical under these circumstances.

We further explored Ca²⁺ entry mechanisms linked to NF-ATc4 activation using direct depolarization of hippocampal neurons (Fig. 2c). A 3-min exposure to 90 mM K⁺ resulted in activation of NF-AT-dependent gene expression. The K⁺-induced increase in NF-AT-dependent transcription was blocked by nifedipine or FK506, but not by AP-5 (Fig. 2c). Pre-exposure to the L-type Ca²⁺ channel agonist (-)-Bay K8644 (5 μ M) strongly augmented the NF-AT transcriptional response to K⁺ depolarization. In contrast to the clear participation of L-type channels, blockade of N-type Ca²⁺ channels by ω -CTx-GVIA (1 μ M) or P/Q- and N-type channels with ω -CTx-MVIIIC (5 μ M) did not reduce the transcriptional response (Fig. 2c). Thus, the K⁺ depolarization acts specifically through L-type Ca²⁺ channels to induce NF-AT-dependent transcription. During the course of spontaneous electrical activity (Fig. 2b), activation of L-type channels is driven by synaptic transmission through NMDA and other glutamate receptors.

In lymphocytes, Ca²⁺ entry regulates the activity of calcineurin, a Ca²⁺/calmodulin-dependent phosphatase that dephosphorylates

NF-ATc family members^{9,12}, unmasking their nuclear localization sequences and leading to nuclear import¹³. To determine whether nuclear translocation of NF-ATc4 takes place in neurons, we monitored cellular localization of enhanced green fluorescent protein (EGFP)-NF-ATc4 with fluorescence imaging (Fig. 3). K⁺ depolarizations produced a rapid redistribution of EGFP-NF-ATc4, monitored with real time imaging of GFP (data not shown). Nuclear NF-ATc4 levels were greatly increased 15 min after the 3-min depolarization and returned to baseline after 2 h (Fig. 3a, b). The persistence of NF-ATc4 in the nucleus of hippocampal neurons following a brief Ca²⁺ pulse is in sharp contrast to lymphocytes, where a sustained elevation of intracellular Ca²⁺ is required to maintain NF-ATc proteins in the nucleus¹⁴. This difference in kinetics might be fundamental to the requirement for rapid responses to Ca²⁺ in neurons. The increase in nuclear-to-cytoplasmic ratio was blocked by the L-type channel antagonist nifedipine and by inhibition of calcineurin with FK506/CsA (Fig. 3a, b). Depolarization in the presence of Bay K8644 produced no greater increase in nuclear-to-cytoplasmic ratio but retarded exit from the nucleus over the 2 h following stimulation (Fig. 3a, b). This altered time course is consistent with the powerful effect of the L-type channel agonist on reporter gene activation (Fig. 2c). Electrical field stimulation (5 Hz for 3 min) and spontaneous electrical activity also induced a strong increase in the level of nuclear NF-ATc4 (Fig. 3c, d), consistent with the effect of electrical activity on NF-AT-dependent transcription (Fig. 2a). Direct stimulation of NMDA receptors (10 μ M NMDA for 3 min) also induced nuclear translocation of NF-ATc4 which was dependent on L-type channels (Fig. 3d). Longer NMDA stimulation (15 min) caused more pronounced nuclear translocation of NF-ATc4, but it too was

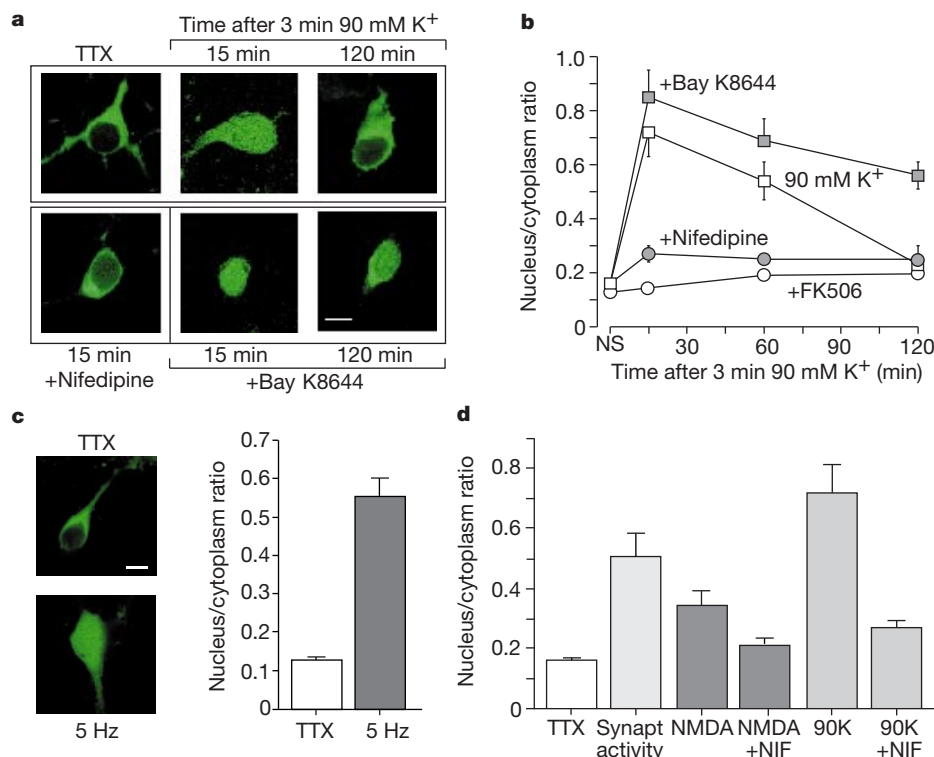


Figure 3 Activity-dependent cytoplasmic-to-nuclear translocation of NF-ATc4 in hippocampal neurons. Cells were transfected with EGFP-NF-ATc4 plasmid (4 μ g) for all imaging experiments. **a**, Confocal sections showing NF-ATc4 distribution following 3 min 90 mM K⁺ depolarization in the absence and presence of nifedipine or Bay K8644. **b**, Confocal quantification of EGFP-NF-ATc4 translocation after 3 min K⁺ stimulation. Nifedipine and FK506/CsA inhibited nuclear import of NF-ATc4, whereas Bay K8644

slowed its nuclear export. **c**, Nuclear translocation of EGFP-NF-ATc4 upon synaptic stimulation. Confocal images and quantification of EGFP-NF-ATc4 translocation upon electrical field stimulation at 5 Hz for 3 min. **d**, Quantification of EGFP-NF-ATc4 translocation in response to spontaneous synaptic activity and direct 3 min NMDA-receptor stimulation and 3 min K⁺ stimulation. Nifedipine inhibited the nuclear translocation of NF-ATc4 induced by NMDA-receptor stimulation.

blocked by nifedipine. These results show that L-type Ca^{2+} channels increase nuclear NF-ATc4, thus activating NF-AT-dependent transcription.

As the rate of nuclear efflux of EGFP-NF-ATc4 in hippocampal neurons could be modulated by L-type channel agonists (Fig. 3a, b), we considered possible export mechanisms. In nonexcitable cells, bidirectional shuttling of NF-ATc proteins between cytoplasm and nucleus is regulated by calcineurin and opposing kinases. Several Ser/Thr kinases, including GSK-3 β ¹⁵, JNK¹⁶, MEKK¹⁷ and casein kinase 1 α ¹⁷ phosphorylate NF-ATc proteins at specific amino-terminal motifs¹³, thereby opposing the actions of calcineurin. The effects of these kinases in neurons were tested by cotransfection with the NF-AT luciferase reporter. Overexpression of GSK-3 β inhibited the activation of NF-AT-dependent transcription in response to PMA plus ionomycin treatment and K^+ depolarization (Fig. 4a), but had no effect on the expression of a constitutively active reporter gene (data not shown). In contrast to GSK-3 β , cotransfection of JNK, constitutively active MEKK or casein kinase 1 α produced no inhibitory effect (Fig. 4b). The physiological importance of endogenous GSK-3 was examined by expression of a dominant negative form of GSK-3 β ¹⁸, which augmented the activation of the NF-AT-dependent reporter gene (Fig. 4c); GSK-3 was present in the nuclear fraction of lysed hippocampal neurons as determined by subcellular fractionation and western blotting (Fig. 4d). Accordingly, we examined the *in vitro* phosphorylation of an NF-ATc4-GST fusion protein by hippocampal extracts and found four distinct phosphorylation bands (Fig. 4e). To test whether any of these were due to endogenous GSK-3, we used antibodies specific for GSK-3 α and β to immunodeplete the extracts. The GSK-3-depleted extracts failed to produce two of the four phosphorylated bands (middle lane, Fig. 4e), in contrast to extracts mock-depleted with protein-G beads, which produced all

four phosphorylation bands (left lane, Fig. 4e). Recognition of substrates by GSK-3 usually requires prior phosphorylation of the substrate by a priming kinase. Using NF-ATc4 that had been treated with GSK-3 α and β -depleted extracts to provide this priming phosphorylation, we found that recombinant GSK-3 β produced two phosphorylation bands with the same mobilities as the bands that had been eliminated by GSK-3 depletion (right lane, Fig. 4e).

We devised a nuclear export assay to assess the effects of GSK-3 β on nuclear export of EGFP-labelled NF-ATc4 in hippocampal neurons. Transfection of neurons with constitutively active calcineurin established a stable starting condition with EGFP-NF-ATc4 abundant in the nucleus (Fig. 4f, g). Nuclear exit of the fluorescently tagged transcription factor was initiated by blocking calcineurin activity with FK506 and CsA, and monitored with measurements of the nuclear-to-cytoplasmic ratio of EGFP fluorescence. Neurons transfected with GSK-3 β showed more than 3-fold faster export than those transfected with empty vector (Fig. 4f, g). Expression of a dominant negative form of GSK-3 β slowed the rate of nuclear export of NF-ATc4 (Fig. 4g), consistent with its effect on the transcriptional response (Fig. 4c). Thus GSK-3 β acts in hippocampal nuclei to phosphorylate NF-ATc4, thereby directing its nuclear export and terminating transcription.

The inositol 1,4,5-triphosphate receptor type 1 (IP₃R1) is found throughout the endoplasmic reticulum network in neurons and is a major determinant of intracellular Ca^{2+} transients. In cerebellar granule cells, transcription of the IP₃R1 gene is regulated by Ca^{2+} entry through L-type channels and is inhibited by FK506 and CsA, implicating calcineurin as a regulator¹⁹. Likewise, in hippocampal neurons, K^+ depolarization in the presence of Bay K8644 caused a substantial increase in IP₃R1 protein levels relative to those observed in neurons stimulated in the presence of FK506 and CsA (Fig. 5a). In

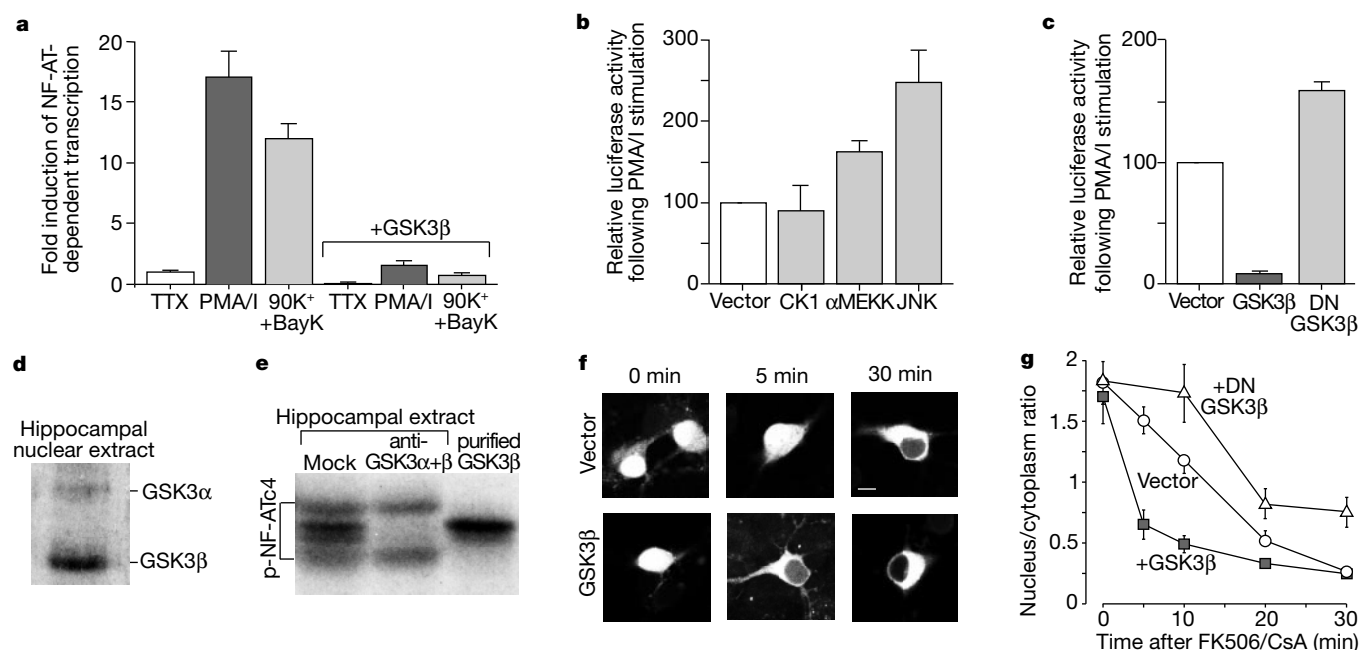


Figure 4 GSK-3 β blocks NF-AT-dependent transcription and promotes nuclear export of NF-ATc4. **a**, Overexpression of GSK-3 β inhibited activation of NF-AT-dependent transcription. Neurons were cotransfected with plasmids for GSK-3 β (4 μg) or empty vector (4 μg), along with NF-AT-luciferase (4 μg) and pBJ5-NF-ATc4 (0.5 μg). **b**, Cotransfection of JNK-1 (4 μg), constitutively active MEKK (4 μg) and casein kinase 1 α (4 μg) produced no inhibitory effect upon NF-AT-dependent luciferase expression. **c**, Cotransfection of dominant negative GSK-3 β (4 μg) augmented the activation of NF-AT-dependent transcription. **d**, Immunoblot showing GSK-3 α and β in the nucleus of

hippocampal neurons. **e**, Endogenous GSK-3 activity in the hippocampus phosphorylated NF-ATc4. **f**, Accelerated nuclear export of NF-ATc4 after overexpression of GSK-3 β , shown in representative confocal sections. Neurons were cotransfected with EGFP-NF-ATc4 (2 μg), constitutively active calcineurin (0.5 μg) and GSK-3 β (4 μg) or empty vector (4 μg). FK506 (400 ng ml^{-1}) and CsA (2 μg ml^{-1}) were added to inhibit calcineurin at 0 min. **g**, Quantification of NF-ATc4 export kinetics in the presence of empty vector (4 μg), GSK-3 β (4 μg) and dominant negative GSK-3 β (4 μg) and dominant negative GSK-3 β (4 μg) after FK506/CsA addition.

preliminary studies, we found that mice with null mutations of NF-ATc4 or mice with null mutations in NF-ATc2 and NF-ATc4 have low levels of expression of the IP₃R1 gene (I.G. and G.C., unpublished data). The 5' flanking region of the murine IP₃R1 gene contains many candidate NF-AT binding sequences, four of which were chosen for further study (Fig. 5b).

Nuclear extracts from rat hippocampi contained binding activity for the NF-AT consensus sequence (CGGGGAAAGTTTGT) from the IP₃R1 5' flanking sequence (Fig. 5c). Binding to this probe was partially competed with an excess of unlabelled oligonucleotide from the distal NF-AT-binding site from the interleukin (IL)-2 gene (ARRE)⁸, and with the unlabelled IP₃R1 site (Fig. 5c). Recombinant NF-ATc4, prepared from transfected Jurkat cells stimulated with PMA and ionomycin (Fig. 5c), produced a DNA-protein complex with the IP₃R1 probe similar to that seen with extracts from hippocampal neurons.

A property of NF-AT transcription complexes^{6,8} is their ability to integrate signalling pathways²⁰ by the combined action of NF-ATc and a nuclear partner protein, NF-ATn²¹. To see whether this was true for the IP₃R1 site, we performed a mixing experiment similar to that used to define the original NF-AT1 transcription complex⁶. By themselves, extracts of nonstimulated hippocampal cells and extracts of PMA-treated JST-cells contained little DNA-binding activity; however, the two extracts in combination produced a substantial enhancement of DNA-binding activity (Fig. 5d). Thus, the IP₃R1 NF-AT consensus sequence has the properties of a functional NF-AT site. We tested whether the nuclear partner protein might be AP-1, as found in lymphocytes¹². There are no consensus AP-1 sites surrounding the neuronal IP₃R1 NF-AT elements, and an unlabelled AP-1-binding oligonucleotide failed to compete with binding of hippocampal extracts to the IP₃R1 NF-

AT site (Fig. 5c, lane 4). This indicates that the nuclear partner of NF-ATc4 in neurons may be a binding protein distinct from AP-1.

The interplay between the regulated nuclear import and export of NF-ATc4 that we have found in hippocampal neurons represents a novel mechanism by which neuronal Ca²⁺ signals may control gene expression (Fig. 5e). Synaptic activity or direct depolarization produced rapid translocation of NF-ATc4 to the nucleus and NF-AT-dependent gene expression, showing that this pathway can function in parallel with other pathways for transcriptional control by Ca²⁺ (refs 22, 23). The regulation of IP₃R1 expression may have interesting implications for forms of synaptic plasticity such as long-term depression, whose induction in the hippocampus and cerebellum depends on Ca²⁺ release from IP₃-sensitive stores^{24–26}. The NF-AT-dependent induction of IP₃R1 could alter the amplitude and/or spatial organization of the Ca²⁺ signal in neurons and would be well suited as a feedback mechanism for stabilization of changes in synaptic strength, which is interesting because transcription and translation are necessary for long-term memory formation^{27,28}. □

Methods

In situ hybridization and RT-PCR

Hybridization with a digoxigenin-labelled riboprobe (Boehringer Mannheim) was performed using standard methods. The NF-ATc4 probe was an *Eco*RI/*Bam*HI fragment containing part of exon 2 of the murine NF-ATc4 gene. RT-PCR, using cDNA prepared from murine and newborn rat hippocampus, was performed according to standard protocols. The primers used were derived from exon 2 and exon 3 of the murine NF-ATc4 gene (5'-GAAGCTACCTCCGGTACAGAG-3' and 5'-GCTTCATAGCTGGCTGTAGCC-3').

Cell culture and transfections

Cultured CA3/CA1 hippocampal neurons were transfected using a Ca²⁺-phosphate procedure at 7 days *in vitro*²⁹. For luciferase assays cell media contained 1–2 μM

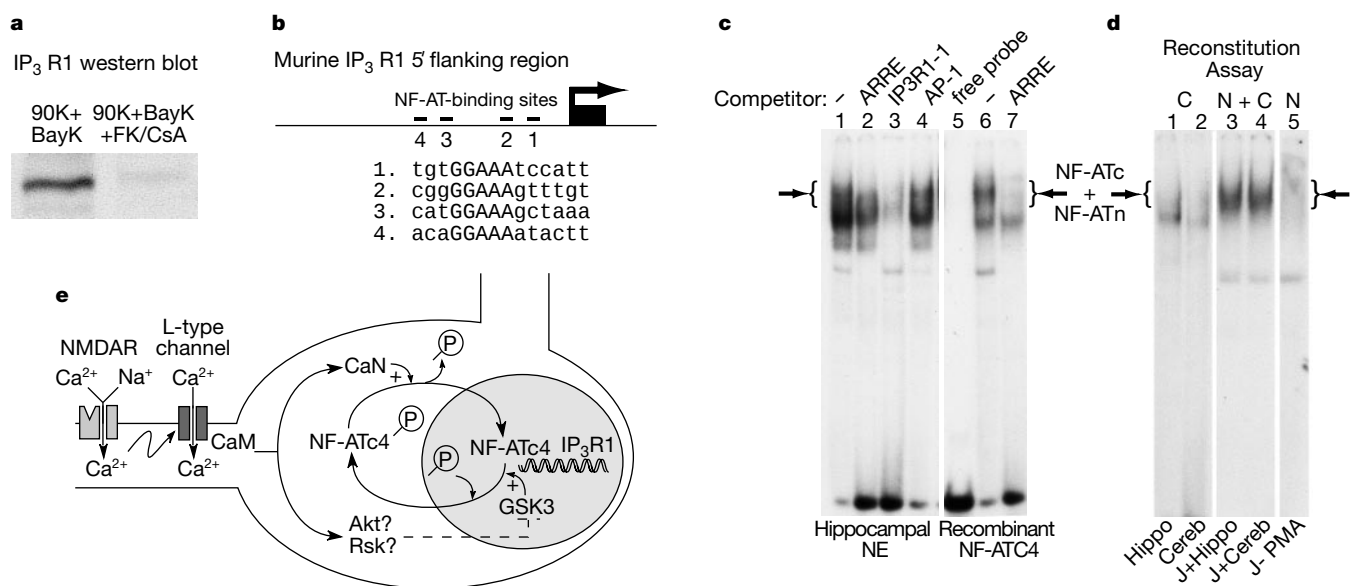


Figure 5 NF-AT-dependent regulation of IP₃R1. **a**, IP₃R1 protein expression is regulated by calcineurin in hippocampal neurons. Western blot showing inhibition by FK506 and CsA of IP₃R1 protein expression induced by L-type channels. **b**, Schematic of the 5' flanking region of the murine IP₃R1 gene. The four indicated NF-AT consensus sites were analysed for NF-AT binding by gel mobility shift assay. **c**, NF-AT in hippocampal nuclear extracts gives rise to specific DNA-binding activity with the IP₃R1 NF-AT consensus site oligonucleotide. A ³²P-labelled oligonucleotide from the putative IP₃R1 NF-AT-binding site 2 was used in gel mobility shift assays with nuclear extracts from rat hippocampi (5 μg) complemented with nuclear extract from PMA-stimulated JST cells (lane 1–4) (2.5 μg). Cold competitor oligonucleotides are indicated and were used at 100-fold excess. Arrows indicate specific NF-AT complexes. Nuclear extracts from NF-ATc4 transfected Jurkat

T-cells stimulated with PMA ionomycin (5 μg) bound to the IP₃R1 probe (lane 6), and the binding was competed by an excess of cold ARRE oligonucleotide (lane 7). Lane 5 shows the free IP₃R1 NF-AT-binding site 2 probe. NE, nuclear extracts. **d**, Cooperative DNA binding to the IP₃R1 probe of NF-ATc from hippocampal and cerebellar extracts and NF-ATn from PMA-stimulated nuclear extracts. Binding reactions were carried out with NF-ATc from hippocampal extracts (10 μg) or cerebellar extracts (10 μg) (lanes 1 and 2), or hippocampal and cerebellar extracts (10 μg) combined with nuclear extract (5 μg) from PMA-stimulated JST cells (lanes 3 and 4), or nuclear extract from PMA stimulated JST-cells (5 μg) alone (lane 5). **e**, Interplay of dephosphorylation and phosphorylation in the regulation of NF-ATc4 in hippocampal neurons.

tetrodotoxin (TTX; Calbiochem) after transfection. Cells were stimulated 48 h after transfection with PMA/1 for 3 h or with 90 mM K⁺ for 3 min. For studies with nifedipine, Bay K8644 or AP-5, cells were pre-incubated for 3 min with the inhibitors before stimulation. The inhibitors were also present during stimulation. For experiments with FK506/CsA, ω -CgTx-GVIA and ω -CTx-MVIIC, the pre-incubation time was 20 min. For assays of spontaneous synaptic activity cells were transfected at 9 days *in vitro* and TTX or pharmacological inhibitors were added 16 h after transfection.

Subcellular fractionation and *in vitro* kinase assays

Nuclear extracts from newborn rat hippocampi were prepared using standard methods and analysed by western blot with a monoclonal antibody specific for GSK-3 α and β (Santa Cruz). Kinase assays were done as described¹⁵ with minor modifications. Samples were analysed by SDS-PAGE and autoradiography. The gel was stained with Coomassie blue to ensure equal loading of the fusion protein.

Plasmids and materials

Detailed description of the plasmids used in these studies can be found at <http://crablab.stanford.edu/>. Pharmacological agents were ionomycin (Calbiochem), phorbol-12-myristate-13-acetate (Calbiochem), tetrodotoxin (Calbiochem), Nifedipine (RBI), D(-)AP-5 (RBI), S(-)BayK 8644 (RBIU), ω -CgTx-GVIA (RBI), ω -CTx-MVIIC (RBI), FK506 (Fujisawa) and cyclosporin A (Sandoz).

Gel mobility shift assays

Nuclear extracts were prepared from newborn rat hippocampi and cerebellum and from NF-ATc4 transfected Jurkat T cells stimulated for 3 h with PMA and ionomycin and from PMA-stimulated JST-B cells as described⁶ with minor modifications for the neuronal extracts. Binding reactions and electrophoretic mobility shift assays were carried out as described⁶. The oligonucleotide sequence of the ³²P-end-labelled oligonucleotide from the putative IP₃R1 NF-AT-binding site 2 from the 5' flanking region of the IP₃R1 gene was 5'-TGACACCCGGGAAAGTTTGTGGAATGAATACGT-3'. The nucleotide sequence from the distal NF-AT binding site from the human IL-2 promoter (ARRE) was 5'-GATCG-GAGGAAAACTGTTTCATACAGAAGGCGT-3' and the AP-1 oligonucleotide sequence was 5'-GGGGTGACTAGGGG-3'.

Western blots for IP₃R1

One-week-old cultured hippocampal neurons were treated with TTX for three days. Stimulations were done as described for the transcriptional assays. Whole cell RIPA lysates were analysed by western blot with a polyclonal antibody specific for IP₃R1, a gift from I. Bezprozvanny.

Received 5 July; accepted 26 August 1999.

- Mulkey, R. M., Endo, S., Shenolikar, S. & Malenka, R. C. Involvement of a calcineurin/inhibitor-1 phosphatase cascade in hippocampal long-term depression. *Nature* **369**, 486–488 (1994).
- Lu, Y. F., Hayashi, Y., Moriwaki, A., Tomizawa, K. & Matsui, H. FK506, a Ca²⁺/calmodulin-dependent phosphatase inhibitor inhibits the induction of long-term potentiation in the rat hippocampus. *Neurosci. Lett.* **205**, 103–106 (1996).
- Winder, D. G., Mansuy, I. M., Osman, M., Moallem, T. M. & Kandel, E. R. Genetic and pharmacological evidence for a novel, intermediate phase of long-term potentiation suppressed by calcineurin. *Cell* **92**, 25–37 (1998).
- Mansuy, I. M., Mayford, M., Jacob, B., Kandel, E. R. & Bach, M. E. Restricted and regulated overexpression reveals calcineurin as a key component in the transition from short-term to long-term memory. *Cell* **92**, 39–49 (1998).
- Klee, C. B., Crouch, T. H. & Krinks, M. H. Calcineurin: a calcium- and calmodulin-binding protein of the nervous system. *Proc. Natl Acad. Sci. USA* **76**, 6270–6273 (1979).
- Flanagan, W. M., Corthesy, B., Bram, R. J. & Crabtree, G. R. Nuclear association of a T-cell transcription factor blocked by FK-506 and cyclosporin A [see comments]. *Nature* **352**, 803–807 (1991).
- Liu, J. *et al.* Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP–FK506 complexes. *Cell* **66**, 807–815 (1991).
- Shaw, J.-P. *et al.* Identification of a putative regulator of early T cell activation genes. *Science* **241**, 202–205 (1988).
- Clipstone, N. A. & Crabtree, G. R. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357**, 695–697 (1992).
- Hoey, T., Sun, Y.-L., Williamson, K. & Xu, X. Isolation of two new members of the NF-AT gene family and functional characterization of the NF-AT proteins. *Immunity* **2**, 461–472 (1995).
- Crabtree, G. R. Generic signals and specific outcomes: signaling through Ca²⁺, calcineurin, and NF-AT. *Cell* **96**, 611–614 (1999).
- Jain, J. *et al.* The T-cell transcription factor NFATp is a substrate for calcineurin and interacts with Fos and Jun. *Nature* **365**, 352–355 (1993).
- Beals, C. R., Clipstone, N. A., Ho, S. N. & Crabtree, G. R. Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. *Genes Dev.* **11**, 824–834 (1997).
- Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. Rapid shuttling of NF-AT in discrimination of Ca²⁺ signals and immunosuppression. *Nature* **383**, 837–840 (1996).
- Beals, C. R., Sheridan, C. M., Turk, C. W., Gardner, P. & Crabtree, G. R. Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* **275**, 1930–1934 (1997).
- Chow, C. W., Rincon, M., Cavanagh, J., Dickens, M. & Davis, R. J. Nuclear accumulation of NFAT4 opposed by the JNK signal transduction pathway. *Science* **278**, 1638–1641 (1997).
- Zhu, J. *et al.* Intramolecular masking of nuclear import signal on NF-AT4 by casein kinase I and MEKK1. *Cell* **93**, 851–861 (1998).
- He, X., Saint-Jeannet, J. P., Woodgett, J. R., Varmus, H. E. & Dawid, I. B. Glycogen synthase kinase-3

- and dorsoventral patterning in *Xenopus* embryos [published erratum appears in *Nature* **375**, 253 (1995)]. *Nature* **374**, 617–622 (1995).
- Genazzani, A. A., Carafoli, E. & Guerini, D. Calcineurin controls inositol 1,4,5-trisphosphate type 1 receptor expression in neurons. *Proc. Natl Acad. Sci. USA* **96**, 5797–5801 (1999).
- Crabtree, G. R. Contingent genetic regulatory events in T lymphocyte activation. *Science* **243**, 355–361 (1989).
- Zhou, P., Sun, L. J., Dotsch, V., Wagner, G. & Verdine, G. L. Solution structure of the core NFATC1/DNA complex. *Cell* **92**, 687–696 (1998).
- Bito, H., Deisseroth, K. & Tsien, R. W. Ca²⁺-dependent regulation in neuronal gene expression. *Curr. Opin. Neurobiol.* **7**, 419–429 (1997).
- Ghosh, A. & Greenberg, M. E. Calcium signaling in neurons: molecular mechanisms and cellular consequences. *Science* **268**, 239–247 (1995).
- Inoue, T., Kato, K., Kohda, K. & Mikoshiba, K. Type 1 inositol 1,4,5-trisphosphate receptor is required for induction of long-term depression in cerebellar Purkinje neurons. *J. Neurosci.* **18**, 5366–5373 (1998).
- Kasano, K. & Hirano, T. Involvement of inositol trisphosphate in cerebellar long-term depression. *NeuroReport* **6**, 569–572 (1995).
- Reyes, M. & Stanton, P. K. Induction of hippocampal long-term depression requires release of Ca²⁺ from separate presynaptic and postsynaptic intracellular stores. *J. Neurosci.* **16**, 5951–5960 (1996).
- Nguyen, P. V., Abel, T. & Kandel, E. R. Requirement of a critical period of transcription for induction of a late phase of LTP. *Science* **265**, 1104–1107 (1994).
- Flexner, L. B., Flexner, J. B. & Roberts, R. B. Memory in mice analyzed with antibiotics. Antibiotics are useful to study stages of memory and to indicate molecular events which sustain memory. *Science* **155**, 1377–1383 (1967).
- Deisseroth, K., Heist, E. K. & Tsien, R. W. Translocation of calmodulin to the nucleus supports CREB phosphorylation in hippocampal neurons. *Nature* **392**, 198–202 (1998).
- Bito, H., Deisseroth, K. & Tsien, R. W. CREB phosphorylation and dephosphorylation: a Ca²⁺- and stimulus duration-dependent switch for hippocampal gene expression. *Cell* **87**, 1203–1214 (1996).

Acknowledgements

We thank T. Hoey for human NF-ATc4 cDNA, I. Bezprozvanny for the IP₃R1 polyclonal antibody, D. Virshup for the casein kinase 1 α expression construct, J. Healy for the MEKK expression construct and M. Karin for the JNK-1 expression construct. We also thank D. Wheeler, G. Pitt, J. H. Bayle and S. H. Park for helpful discussions. K.S. is a Stanford graduate fellow and a Howard Hughes Medical Institute predoctoral fellow. G.R.C. is an Investigator of the Howard Hughes Medical Institute. R.W.T. is a McKnight Senior Investigator. This work was supported by grants from the NIH and the Mathels Charitable Trust.

Correspondence and requests for materials should be addressed to G.R.C. (e-mail: hf.grc@forysythe.stanford.edu).

Two subsets of memory T lymphocytes with distinct homing potentials and effector functions

Federica Sallusto^{*}, Danielle Lenig^{*}, Reinhold Förster[†], Martin Lipp[†] & Antonio Lanzavecchia^{*}

^{*} Basel Institute for Immunology, Grenzacherstrasse 487, Postfach, CH-4005 Basel, Switzerland

[†] Max-Delbrueck-Center for Molecular Medicine, Robert Rossle Strasse 10, 13122 Berlin-Buch, Germany

Naive T lymphocytes travel to T-cell areas of secondary lymphoid organs in search of antigen presented by dendritic cells^{1,2}. Once activated, they proliferate vigorously, generating effector cells that can migrate to B-cell areas or to inflamed tissues^{3–6}. A fraction of primed T lymphocytes persists as circulating memory cells that can confer protection and give, upon secondary challenge, a qualitatively different and quantitatively enhanced response^{7–9}. The nature of the cells that mediate the different facets of immunological memory remains unresolved. Here we show that expression of CCR7, a chemokine receptor that controls homing to secondary lymphoid organs, divides human memory T cells into two functionally distinct subsets. CCR7⁺ memory cells express receptors for migration to inflamed tissues and display

immediate effector function. In contrast, $CCR7^+$ memory cells express lymph-node homing receptors and lack immediate effector function, but efficiently stimulate dendritic cells and differentiate into $CCR7^-$ effector cells upon secondary stimulation. The $CCR7^+$ and $CCR7^-$ T cells, which we have named central memory (T_{CM}) and effector memory (T_{EM}), differentiate in a step-wise fashion from naive T cells, persist for years after immunization and allow a division of labour in the memory response.

When blood-borne naive T cells home to lymph nodes, they first roll on high endothelial venules using CD62L. This allows the chemokine receptor CCR7 to engage its ligand SLC, which is displayed by endothelial cells¹⁰. The CCR7–SLC interaction activates integrins that promote firm adhesion and transmigration of the T cells into the lymph node^{11,12}. In contrast to naive T cells, memory/effector cells migrate mostly through peripheral tissues¹³.

This migration mediates rapid protective responses and is controlled by the expression of different sets of integrins and chemokine receptors^{1,14}. However, some memory T cells must also reach the lymph nodes to mount secondary proliferative responses. We considered whether the two facets of the memory response might depend on subsets of memory T cells invested with distinct homing and effector capacities.

Because CCR7 and CD62L are essential for lymphocyte migration to lymph nodes^{1,15}, the co-expression of these receptors might distinguish a putative subset of memory T cells that home to lymph nodes. Human naive and memory T cells can be identified by the reciprocal expression of the CD45RA or CD45RO isoforms¹⁶. Staining of peripheral blood T cells with antibodies to CD45RA and CCR7 revealed three subsets of $CD4^+$ cells: one naive $CD45RA^+CCR7^-$; and two memory subsets, $CD45RA^-CCR7^+$ and

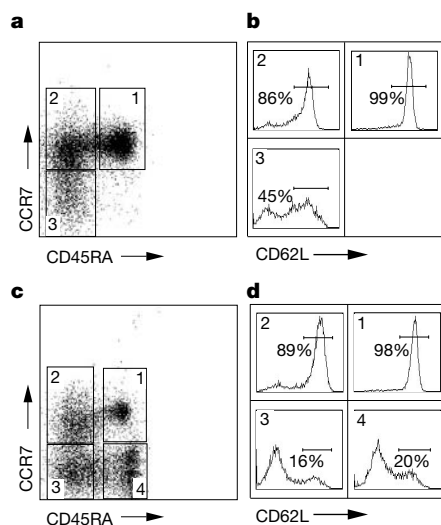


Figure 1 CCR7 and CD62L are co-expressed on a subset of peripheral blood memory $CD4^+$ and $CD8^+$ T cells. $CD4^+$ (a, b) and $CD8^+$ (c, d) lymphocytes were stained with monoclonal antibodies to CD45RA and CCR7, which identified three and four subsets, respectively. These subsets were sorted and analysed for the expression of CD62L, and the percentage of bright cells is indicated (b, d). Upon serial analysis, the proportion of cells in the different compartments was rather stable in the same individual, but more variable among individuals, the variability being more pronounced in the CD8 than in the CD4 compartment. Comparable results were obtained using two anti-CCR7 antibodies (clones 3D12 and 10H5).

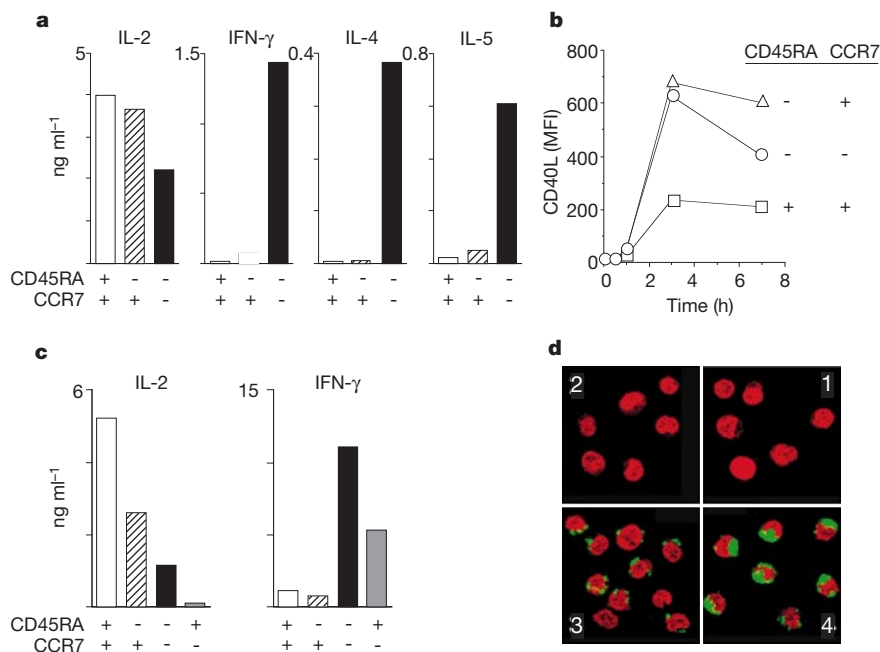


Figure 2 $CCR7^+$ and $CCR7^-$ memory T cells display different effector functions. a, b, The three subsets of $CD4^+$ T cells were sorted according to the expression of CCR7 and CD45RA as in Fig. 1 and tested for their capacity to produce IL-2, IFN- γ , IL-4 and IL-5 (a) and for the kinetics of surface CD40L upregulation (b) following polyclonal stimulation. c, d, The four subsets of $CD8^+$ T cells were sorted according to the expression of CCR7

and CD45RA as in Fig. 1 and tested for their capacity to produce IL-2 or IFN- γ (c) or were immediately stained with anti-perforin antibody (green) and counterstained with propidium iodide (red) (d). In the $CD8^+ CD45RA^+$ compartment, CCR7 expression allows us to discriminate naive cells (1) from effector cells (4) (ref. 26). Comparable results were obtained in 12 healthy donors.

CD45RA⁺CCR7⁺ (Fig. 1a). Both naive and CCR7⁺ memory cells expressed high levels of CD62L, whereas the CCR7⁻ memory cells expressed CD62L to a lower and variable extent (Fig. 1b). Within CD8⁺ T cells, the same three subsets could be identified, with an extra subset of CD45RA⁺CCR7⁻ cells (Fig. 1c). In addition, the two CCR7⁺ subsets expressed high levels of CD62L, whereas most of the cells among the two CCR7⁻ subsets lacked CD62L (Fig. 1d). This shows that lymph-node-homing receptors are expressed on a distinct subset of memory CD4 and CD8 T cells; furthermore, it identifies a subset of CD8⁺ cells that is CD45RA⁺, but that lacks both CCR7 and CD62L.

A number of chemokine receptors and adhesion molecules are involved in lymphocyte migration to secondary lymphoid organs or to tissues under homeostatic or inflammatory conditions^{1,14}. CCR7⁺ memory T cells express high levels of β 1 and β 2 integrins, which are required for homing to inflamed tissues¹⁷, as well as tissue-specific homing receptors such as CD103 and CLA (Table 1). Receptors for inflammatory chemokines, such as CCR1, CCR3 and CCR5, which have been found on memory/effector cells^{18,19}, were selectively expressed in the CCR7⁺ memory subset. On the other hand, CCR7⁺ memory cells had a distinct phenotype. They expressed intermediate levels of β 1 and β 2 integrins, as well as CCR4, CCR6 and CXCR3 on various proportions of cells. T-cell activation markers such as CD69 and CD25 were expressed only on a small fraction of the two memory subsets, whereas HLA-DR was expressed on about 10% of the CCR7⁺ memory cells. Together, these results imply that CCR7⁺ memory T cells share migratory routes with naive T cells, although the expression of additional receptors such as CCR4 may allow them to respond to a wider spectrum of chemokines and interact more effectively with dendritic cells^{20,21}.

Memory T cells carrying distinct homing receptors might participate in different types of immune responses and therefore might have different effector capacities. T-cell help for dendritic cells and B cells is dependent on expression of CD40L²², whereas protective responses in the tissues are mediated by T cells that produce effector cytokines, such as interferon- γ (IFN- γ) or interleukin-4 (IL-4), or release stored perforin^{23,24}. The naive and two memory CD4 subsets were sorted and compared for their capacity to produce cytokines and upregulate CD40L following stimulation. As shown in Fig. 2a, both naive T cells and CCR7⁺ memory cells produced IL-2 only. In contrast, the CCR7⁺ memory subset produced high levels of IL-4, IL-5 and IFN- γ and moderately reduced levels of IL-2. Upon activation, the extent of CD40L upregulation was comparable in the two memory subsets and was higher than in naive T cells; however, the kinetics of upregulation were comparable, indicating that, unlike tonsil T cells²⁵, circulating memory T cells do not contain stored CD40L (Fig. 2b). Rapid production of IFN- γ was detected in most CCR7⁺, but only a negligible fraction of CCR7⁺

memory cells following stimulation with autologous dendritic cells pulsed with a bacterial superantigen (Fig. 3). Within the four CD8⁺ T-cell subsets, IFN- γ production and perforin-containing granules were restricted to the CCR7⁺ cells (Fig. 2c, d). Perforin expression was particularly prominent in the CD45RA⁺CCR7⁺ population that corresponds to a reported population of terminally differentiated CD27⁺ effector T cells²⁶. CD8⁺ cells progressively lost IL-2 production, presumably as a function of their differentiation from naive cells to effectors.

We then examined the activation requirements, which are less stringent for memory cells⁹. Compared with naive cells, both memory subsets displayed increased sensitivity to stimulation by anti-CD3, both in the presence and in the absence of costimulation, although CCR7⁺ cells were consistently more responsive (Fig. 4a). The capacity to stimulate IL-12 production was tested by co-culturing T cells with dendritic cells pulsed with different doses of the bacterial superantigen TSST. CCR7⁺ memory cells efficiently stimulated IL-12 production by dendritic cells at both high and low doses of TSST, whereas naive T cells were much less effective and only stimulated IL-12 at the highest TSST dose (Fig. 4b). Thus, because of their lower triggering threshold and higher capacity to upregulate CD40L, CCR7⁺ memory cells can function as potent activators of dendritic cells.

The above results indicated that two subsets of circulating memory T cells with different functional capacities can be discriminated by the expression of CCR7; therefore, we considered whether memory for antigens would be present in both subsets. As reported¹⁶, proliferative responses to tetanus toxoid can not be detected in naive T cells; however, they were consistently found in both the CCR7⁺ and CCR7⁻ memory subsets, even ten years after vaccination (Fig. 5). Following a booster with tetanus toxoid, the proliferative responses increased in both subsets, indicating that

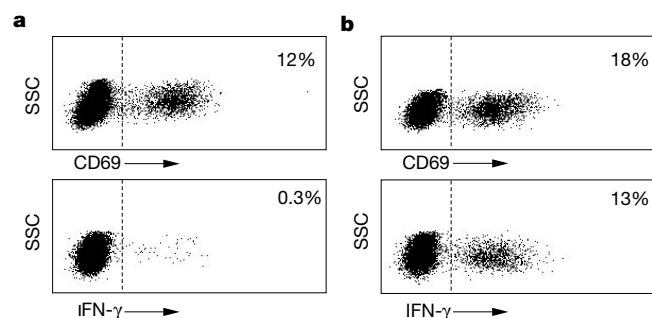


Figure 3 Rapid production of IFN- γ following stimulation of CCR7⁺ memory T cells. CD45RA⁺CCR7⁺ (a) and CD45RA⁺CCR7⁻ (b) CD4 T cells were stimulated for 7 h with autologous dendritic cells pulsed with 100 ng ml⁻¹ TSST and stained with antibodies to CD69 and IFN- γ . CD69⁺ cells were less than 2% in unstimulated cultures.

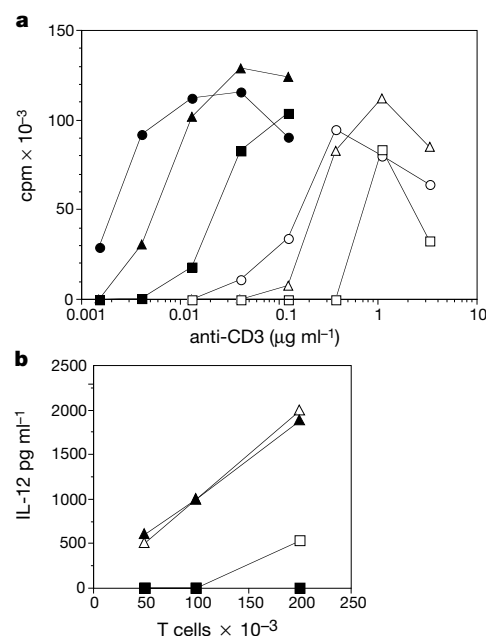


Figure 4 CCR7⁺ memory cells show enhanced responsiveness to T-cell receptor triggering and potently activate dendritic cells to produce IL-12. **a**, Proliferative response of naive T cells (squares), CCR7⁺ (triangles) and CCR7⁻ (circles) memory T cells to different concentrations of plastic-bound anti-CD3 monoclonal antibody in the absence (empty symbols) or in the presence (filled symbols) of anti-CD28. **b**, IL-12 p70 production by dendritic cells cultured with naive T cells (squares) or CCR7⁺ memory T cells (triangles). Dendritic cells were pulsed with toxic shock syndrome toxin (TSST) at 100 ng ml⁻¹ (empty symbols) or 1 ng ml⁻¹ (filled symbols). Both T-cell populations contained similar proportions of V β 2⁺ cells.

Table 1 Surface molecules on peripheral blood naive and memory CD4⁺ T-cell subsets

		CD45RA ⁺ CCR7 ⁺	CD45RA ⁻ CCR7 ⁺	CD45RA ⁻ CCR7 ⁻
CD3	(%)	>99	>99	>99
CD69	(%)	<0.1	2	3
CD25	(%)	<0.1	4	8
HLA-DR	(%)	<0.1	1	12
CD18	(MFI)	38	49	75
CD11a	(MFI)	90	140	218
CD11b	(%)	<0.1	<0.1	35
CD29	(MFI)	0	10/36*	43
CD49d	(MFI)	10	19/2	33/2
CD49e	(%)	<0.1	<0.1	10
CLA	(%)	<0.1	4	25
CD103	(%)	<0.1	<0.1	1
CXCR4	(%)	98	22	11
CCR4	(%)	<0.1	18	6
CCR6	(%)	<0.1	40	45
CXCR3	(%)	<0.1	34	61
CCR1	(%)	<0.1	1	14
CCR3	(%)	<0.1	<0.1	4
CCR5	(%)	<0.1	2	52

Three subsets of peripheral blood CD4⁺ cells were sorted by expression of CD45RA or CD45RO and CCR7. The sorted cells were stained and analysed for the expression of adhesion molecules and chemokine receptors. MFI, mean fluorescence intensity.
* Mean value of major and minor peak.

their relative proportions are not affected by recent antigenic stimulation. In all cases, IFN- γ was produced by only the CCR7⁻ memory cells. Comparable results were obtained by analysing the response to tetanus toxoid or hepatitis B surface antigen in five primed donors, showing that both memory subsets contain clonally expanded antigen-specific T cells, but these differ in their effector capacity. The presence of expanded antigen-specific memory cells in both subsets at different times after antigenic stimulation indicates that homeostatic mechanisms may maintain cells in both compartments^{16,27,28}.

What is the relationship between naive cells and the cells in the two memory subsets? When peripheral blood naive T cells were polyclonally stimulated, all cells became CD45RO⁺ after 10 days (data not shown), but most of the cells retained CCR7 expression, whereas only a few acquired the capacity to produce IL-4 or IFN- γ (Fig. 6a–c). When the CCR7⁺ and CCR7⁻ cells were sorted and stimulated, IL-4, IL-5 and IFN- γ were found to be exclusively

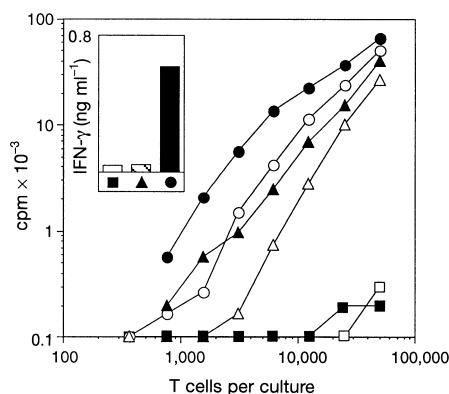


Figure 5 Proliferative responses to recall antigens can be detected in both CCR7⁺ and CCR7⁻ subsets but not in naive T cells. Proliferative response of CD45RA⁺ CCR7⁺ naive T cells (squares), CD45RA⁻ CCR7⁺ (triangles) and CD45RA⁻ CCR7⁻ memory T cells (circles) in response to tetanus toxoid presented by autologous monocytes. Responder cells from the same individual were tested 10 years after vaccination (empty symbols) and two weeks after a booster (filled symbols). Inset shows IFN- γ production in the 24-h culture supernatant using an input of memory cells giving comparable proliferative responses.

produced by the CCR7⁻ cells (Fig. 6d). Thus, with respect to the parameters analysed, it appears that the same functional subsets of memory T cells that are detected *in vivo* can be generated by stimulation of naive cells in short-term culture.

When peripheral blood CCR7⁺ memory T cells were sorted and stimulated under the same conditions, almost all the cells that were recovered after 10 days had lost CCR7 expression and had acquired the capacity to produce effector cytokines upon further stimulation (Fig. 6e–g), indicating that this subset is poised to generate effector cells. Finally, peripheral blood CCR7⁻ memory cells, after stimulation and expansion, retained their CCR7⁻ phenotype and effector function (Fig. 6h–j). This indicates that, at least *in vitro*, there may be a stepwise differentiation from naive T cells to CCR7⁺ memory to CCR7⁻ memory/effector T cells. This possibility is supported by analysis of the telomere length, which decreases as a function of cell division²⁹. As shown in Fig. 6k, the length of telomeres in peripheral blood CD4 subsets decreased progressively from naive to memory

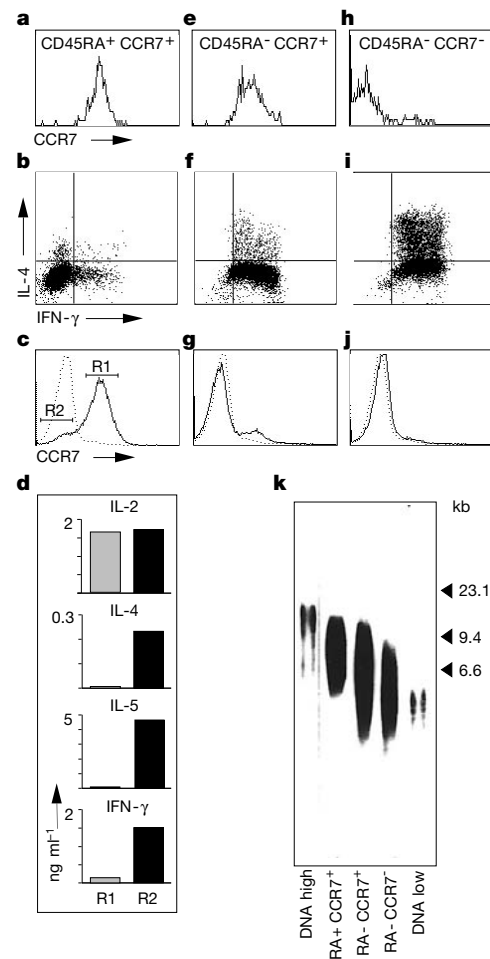


Figure 6 Differentiation potential of naive and memory T-cell subsets. **a–d**, Loss of CCR7 following *in vitro* stimulation of naive T cells correlates with acquisition of effector function. CD4⁺ naive T cells (CD45RA⁺, CCR7⁺) were sorted from peripheral blood (**a**), stimulated with anti-CD3 + anti-CD28, expanded for 10 days in IL-2 and tested for their capacity to produce IFN- γ and IL-4 (**b**) or for CCR7 expression (**c**). CCR7⁺ (R1) and CCR7⁻ cells (R2) were sorted and immediately tested for their capacity to produce cytokines following polyclonal stimulation (**d**). **e–g**, Rapid polarization of CCR7⁺ memory T cells following *in vitro* stimulation. CD4⁺, CD45RA⁻, CCR7⁺ T cells were sorted from peripheral blood (**e**), stimulated, expanded and tested for cytokine production (**f**) and CCR7 expression (**g**). **h–j**, CD4⁺, CD45RA⁻, CCR7⁻ T cells isolated and stimulated as above retained a stable effector phenotype. **k**, Length of telomeres in peripheral blood naive and memory CD4⁺ T-cell subsets. kb, kilobases.

cells, but was consistently higher in the CCR7⁺ than in the CCR7⁻ subset, suggesting that the latter had undergone a larger number of divisions.

We have shown that immunological memory is displayed by distinct T-cell subsets: lymph-node-homing cells lacking inflammatory and cytotoxic function (which we define as central memory T cells, T_{CM}) and tissue-homing cells endowed with various effector functions (which we define as effector memory T cells, T_{EM}). These two subsets allow a division of labour among memory cells. On the one hand, T_{EM} cells represent a readily available pool of antigen-primed cells which can enter peripheral tissues to mediate inflammatory reactions or cytotoxicity, thus rapidly containing invasive pathogens. On the other hand, the newly described T_{CM} cells represent a clonally expanded antigen-primed population which travels to secondary lymphoid organs and, upon a secondary challenge, can efficiently stimulate dendritic cells, help B cells and generate a new wave of effector cells.

Our results indicate a precursor-product relationship between the two memory subsets. *In vitro* stimulation of naive T cells results in the generation of both T_{CM} and T_{EM} cells, whereas stimulation of T_{CM} cells results in their efficient differentiation to T_{EM} cells. Furthermore, antigen-specific T_{CM} and T_{EM} cells persist *in vivo* for up to ten years and their relative proportions do not change after a booster immunization. These data are consistent with a linear differentiation model in which naive T cells differentiate first to T_{CM} and then to T_{EM} cells, depending on the strength and duration of T-cell receptor stimulation and the presence or absence of polarizing cytokines²¹. Understanding the mechanisms that generate and maintain the two types of memory cells will help to manipulate immunological memory for vaccination and for adoptive immunotherapy. □

Methods

Sorting and FACS analysis

Peripheral blood mononuclear cells were stained with a rat monoclonal antibody (mAb) specific for CCR7 (3D12, IgG2a) followed by a fluorescein isothiocyanate (FITC)-labelled mouse anti-rat IgG2a mAb (PharMingen) or alternatively by a phycoerythrin (PE)-labelled goat anti-rat immunoglobulin polyclonal antiserum (Southern Biotechnology Associates). The 3D12 mAb completely inhibited migration of peripheral blood T cells in response to secondary lymphoid tissue chemokine (SLC) and EB1-ligand chemokine (ELC) (R.F., unpublished data) and did not affect the response of T cells to mitogenic or antigenic stimulation (E.S., unpublished data). In addition, 3D12 stained all cell lines that expressed CCR7 messenger RNA, but did not stain CCR7 mRNA-negative cells. In some experiments, a mouse mAb specific for CCR7 (10H5, IgG3; produced by L. Wu, LeukoSite) was used with comparable results. The following PE-, PC5- or APC-labelled mouse mAbs were used in different combinations: anti-CD45RA (ALB11, IgG1); anti-CD45R0 (UCHL1, IgG2a); anti-CD3 (UCHT1, IgG1); anti-CD4 (13B8.2, IgG1); anti-CD8 (B9.11, IgG1); anti-CD11a (25.3, IgG1); anti-CD11b (Bear1, IgG1); anti-CD18 (7E4, IgG1); anti-CD49d (HP2/1, IgG1); anti-CD49e (SAM1, IgG2b); anti-CD29 (K20, IgG2a); anti-CD103 (2G5, IgG2a); anti-CD69 (TP1.55, IgG2b); anti-CD25 (B1.49, IgG2a); anti-HLA-DR (B8.12, IgG2b); anti-CD40L (TRAP-1, IgG1) (all from Immunotech); and anti-CLA (HECA-205, rat IgG1; PharMingen). Staining for chemokine receptors was carried out using the following mouse mAbs (all produced at LeukoSite): anti-CCR1 (2D4, IgG1); anti-CCR3 (7B11, IgG2a); anti-CCR4 (1G1, IgG1); anti-CCR5 (2D7, IgG2a); anti-CCR6 (11A9, IgG1); anti-CXCR3 (1C6, IgG1); and anti-CXCR4 (12G5, IgG2a). Cells were sorted using a fluorescence-activated cell sorter (FACS Vantage) and analysed on a FACScalibur (Becton Dickinson Systems). Sorted cells were immobilized on poly-L-lysine coated slides, fixed in 2% paraformaldehyde and permeabilized in 0.1% Triton-X100 before intracellular staining with an anti-perforin mAb (284, IgG2b; PharMingen), followed by FITC-labelled goat anti-mouse immunoglobulin and propidium iodide to visualize the nuclei by confocal microscopy. The length of telomeres was determined using a Teloquant kit (PharMingen).

Cytokine detection

T cells were stimulated with 10 µg ml⁻¹ anti-CD3 antibody (TR66, IgG1) and 10⁻⁷ M phorbol 12-myristate 13-acetate (PMA; Sigma). Cytokine production was measured in the 24-h culture supernatants by ELISA using matched pairs of antibodies specific for IL-2, IL-4, IL-5, IFN-γ (PharMingen). For cytokine detection at the single-cell level, T cells were stimulated with 10⁻⁷ M PMA and 1 µg ml⁻¹ ionomycin for 4 h, or with autologous dendritic cells pulsed with 100 ng ml⁻¹ TSSST for 7 h in 10 µg ml⁻¹ brefeldin A. Cells were fixed and permeabilized with PBS containing FCS (2%) and saponin (0.5%) and stained with FITC-labelled anti-IFN-γ (IgG1) and PE-labelled anti-IL-4 (IgG2b) or PE-labelled anti-CD69 mAbs.

Cell cultures

Sorted cells were stimulated with plastic-bound anti-CD3 (TR66) and anti-CD28 (CD28.1, IgG1; provided by D. Olive) in RPMI 1640 medium containing 2 mM L-glutamine, 1% non-essential amino acids, 1% sodium pyruvate, 50 µg ml⁻¹ kanamycin, (complete medium; Gibco BRL) and 10% FCS (HyClone Laboratories). The cells were expanded with 500 U ml⁻¹ IL-2. Monocytes were isolated using magnetic beads coated with anti-CD14 mAb (Miltenyi), irradiated (3,000 rad), pulsed with 5 µg ml⁻¹ TT and cultured with different numbers of T cells. IFN-γ was measured by ELISA in the 24-h culture supernatant. ³H-thymidine incorporation was measured on day five. Immature dendritic cells were produced by culturing peripheral blood CD14⁺ monocytes with GM-CSF + IL-4 for six days. The cells were pulsed with different concentrations of TSSST and cultured with graded numbers of T cells. IL-12 p70 was measured by ELISA (PharMingen) in the 24-h culture supernatant.

Received 6 May; accepted 4 August 1999.

- Butcher, E. C. & Picker, L. J. Lymphocyte homing and homeostasis. *Science* **272**, 60–66 (1996).
- Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998).
- MacLennan, I. C. *et al.* The changing preference of T and B cells for partners as T-dependent antibody responses develop. *Immunol. Rev.* **156**, 53–66 (1997).
- Garside, P. *et al.* Visualization of specific B and T lymphocyte interactions in the lymph node. *Science* **281**, 96–99 (1998).
- Mackay, C. R. Homing of naive, memory and effector lymphocytes. *Curr. Opin. Immunol.* **5**, 423–427 (1993).
- Austrup, F. *et al.* P- and E-selectin mediate recruitment of T-helper-1 but not T-helper-2 cells into inflamed tissues. *Nature* **385**, 81–83 (1997).
- Ahmed, R. & Gray, D. Immunological memory and protective immunity: understanding their relationship. *Science* **272**, 54–60 (1996).
- Zinkernagel, R. M. *et al.* On immunological memory. *Annu. Rev. Immunol.* **14**, 333–367 (1996).
- Dutton, R. W., Bradley, L. M. & Swain, S. L. T cell memory. *Annu. Rev. Immunol.* **16**, 201–223 (1998).
- Gunn, M. D. *et al.* A chemokine expressed in lymphoid high endothelial venules promotes the adhesion and chemotaxis of naive T lymphocytes. *Proc. Natl Acad. Sci. USA* **95**, 258–263 (1998).
- Campbell, J. J. *et al.* Chemokines and the arrest of lymphocytes rolling under flow conditions. *Science* **279**, 381–384 (1998).
- Campbell, J. J. *et al.* 6-C-kine (SLC), a lymphocyte adhesion-triggering chemokine expressed by high endothelium, is an agonist for the MIP-3β receptor CCR7. *J. Cell Biol.* **141**, 1053–1059 (1998).
- Mackay, C. R., Marston, W. L. & Dudler, L. Naive and memory T cells show distinct pathways of lymphocyte recirculation. *J. Exp. Med.* **171**, 801–817 (1990).
- Baggiolini, M. Chemokines and leukocyte traffic. *Nature* **392**, 565–568 (1998).
- Gunn, M. D. *et al.* Mice lacking expression of secondary lymphoid organ chemokine have defects in lymphocyte homing and dendritic cell localization. *J. Exp. Med.* **189**, 451–460 (1999).
- Michie, C. A., McLean, A., Alcock, C. & Beverley, P. C. Lifespan of human lymphocyte subsets defined by CD45 isoforms. *Nature* **360**, 264–265 (1992).
- Baron, J. L., Madri, J. A., Ruddle, N. H., Hashim, G. & Janeway, C. A. Jr Surface expression of alpha 4 integrin by CD4 T cells is required for their entry into brain parenchyma. *J. Exp. Med.* **177**, 57–68 (1993).
- Wu, L. *et al.* CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, *in vitro*. *J. Exp. Med.* **185**, 1681–1691 (1997).
- Sallusto, F., Mackay, C. R. & Lanzavecchia, A. Selective expression of the eotaxin receptor CCR3 by human T helper 2 cells. *Science* **277**, 2005–2007 (1997).
- Tang, H. L. & Cyster, J. G. Chemokine up-regulation and activated T cell attraction by maturing dendritic cells. *Science* **284**, 819–822 (1999).
- Sallusto, F., Lenig, D., Mackay, C. R. & Lanzavecchia, A. Flexible programs of chemokine receptor expression on human polarized T helper 1 and 2 lymphocytes. *J. Exp. Med.* **187**, 875–883 (1998).
- Grewal, I. S. & Flavell, R. A. CD40 and CD154 in cell-mediated immunity. *Annu. Rev. Immunol.* **16**, 111–135 (1998).
- Abbas, A. K., Murphy, K. M. & Sher, A. Functional diversity of helper T lymphocytes. *Nature* **383**, 787–793 (1996).
- Kagi, D., Ledermann, B., Burki, K., Zinkernagel, R. M. & Hengartner, H. Molecular mechanisms of lymphocyte-mediated cytotoxicity and their role in immunological protection and pathogenesis *in vivo*. *Annu. Rev. Immunol.* **14**, 207–232 (1996).
- Casamayor-Palleja, M., Khan, M. & MacLennan, I. C. A subset of CD4⁺ memory T cells contains preformed CD40 ligand that is rapidly but transiently expressed on their surface after activation through the T cell receptor complex. *J. Exp. Med.* **181**, 1293–1301 (1995).
- Hamann, D. *et al.* Phenotypic and functional separation of memory and effector human CD8⁺ T cells. *J. Exp. Med.* **186**, 1407–1418 (1997).
- Sprent, J., Tough, D. F. & Sun, S. Factors controlling the turnover of T memory cells. *Immunol. Rev.* **156**, 79–85 (1997).
- Tanchot, C. & Rocha, B. The organization of mature T-cell pools. *Immunol. Today* **19**, 575–579 (1998).
- Weng, N. P., Hathcock, K. S. & Hodes, R. J. Regulation of telomere length and telomerase in T and B cells: a mechanism for maintaining replicative potential. *Immunity* **9**, 151–157 (1998).

Acknowledgements

We thank K. Hannestad and K. Karjalainen for critical reading; J. C. Howard for his help as 'wordsmith'; M. Dessing and A. Pickert for cell sorting; A. Hoy for help in the initial experiments; LeukoSite Inc., Cambridge, Massachusetts for providing antibodies to chemokine receptors; and L. Wu for providing the CCR7 and CCR4 antibodies before publication. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche Ltd, Basel, Switzerland.

Correspondence should be addressed to F.S. (e-mail: sallusto@bii.ch).

Atomic structure of the GCSF–receptor complex showing a new cytokine–receptor recognition scheme

Masaharu Aritomi^{*†}, Naoki Kunishima^{*}, Tomoyuki Okamoto^{*‡}, Ryota Kuroki[‡], Yoshimi Ota^{*†} & Kosuke Morikawa^{*}

^{*} Biomolecular Engineering Research Institute, 6-2-3 Furuedai, Suita-city, Osaka 565-0874, Japan

[‡] Central Laboratories for Key Technology, Kirin Brewery Co. Ltd, 13-5 Fukuura, Kanazawa-ku, Yokohama-city, Kanagawa 236-0004, Japan

[†] Present addresses: Asahi Chemical Industry Co. Ltd, 2-1 Samejima, Fuji-city, Shizuoka 416-8501, Japan (M.A.); Toray Industrial, Inc., 1111 Tebiri, Kamakura-city, Kanagawa 248-0036, Japan (Y.O.).

Granulocyte colony-stimulating factor (GCSF) is the principal growth factor regulating the maturation, proliferation and differentiation of the precursor cells of neutrophilic granulocytes¹ and is used to treat neutropenia². GCSF is a member of the long-chain subtype of the class 1 cytokine superfamily, which includes growth hormone, erythropoietin, interleukin 6 and oncostatin M (ref. 3). Here we have determined the crystal structure of GCSF complexed to the BN–BC domains, the principal ligand-binding region of the GCSF receptor (GCSFR). The two receptor domains form a complex in a 2:2 ratio with the ligand, with a non-crystallographic pseudo-twofold axis through primarily the inter-domain region and secondarily the BC domain. This structural view of a gp130-type receptor–ligand complex presents a new molecular basis for cytokine–receptor recognition.

The mature GCSFR consists of an immunoglobulin-like domain, a cytokine receptor homologous (CRH) region, gs-CRH, three fibronectin type III (FNIII)-like domains, a transmembrane domain and a cytoplasmic domain³ (Fig. 1a). Conserved in class 1 cytokine receptors, the gs-CRH region is the primary interface for GCSF binding and *in vivo* activation, with the FNIII-like and

immunoglobulin-like domains involved at a more subsidiary level⁴. This region is further divided into amino-terminal (BN) and carboxy-terminal (BC) domains. As GCSFR share both domain organization and primary sequence with gp130, a signal transducer for interleukin (IL)-6 family proteins, it can be classified into the gp130-type receptor family.

The complex stoichiometry of GCSF and the receptor remains controversial, as GCSF and different permutations of the GCSFR extracellular domains form 1:1, 2:2, and/or 4:4 complexes, depending on the experiment^{5–9}. However, the binding scheme is clearly different from the 1:2 ligand–receptor stoichiometry found in complexes with growth hormone (GH)¹⁰ and erythropoietin (EPO)¹¹.

The crystallographic study of GCSF¹² and the NMR study of the BC domain¹³ of the receptor revealed topological similarities to GH and to the GH receptor (GHR) and the EPO receptor (EPOR), respectively. In addition, the crystal structure of the ligand-free CRH region from gp130 (ref. 14) showed an L-shaped architecture, as observed in the structures of GH–GHR¹⁰ and EPO–EPOR¹¹. However, a structural view of the recognition between GCSF and GCSFR remained elusive.

The X-ray structure determination of the GCSF–gs-CRH complex at 2.8 Å (Table 1) presented here reveals a recognition scheme distinct from those of the GH–GHR or EPO–EPOR complexes (Fig. 1). The two sets of 1:1 complexes of GCSF–gs-CRH in the asymmetric unit are related by a non-crystallographic pseudo-twofold axis, representing a 2:2 ligand–receptor binding scheme (Fig. 1). Each component adopts the same structural topology determined for the components individually; the root mean square (r.m.s.) deviation for the Cα atoms of the four-helix bundle regions is less than 1.0 Å between the free and complex states, and the BN and BC domains in gs-CRH form an L-shaped structure, as observed in the CRH region of ligand-free gp130 (ref. 14). The molecular arrangement of the ligand and the receptor is similar to that of the IL-4–IL-4R complex¹⁵, but differs from those found in the GH–GHR¹⁰ and EPO–EPOR complexes¹¹, where the ligand is sandwiched between two interfaces and lies outside the elbow of the two L-shaped CRH regions.

gs-CRH binds GCSF through two surfaces; the major interface is formed within the 1:1 complex (Fig. 1d), whereas the minor interface,

Table 1 Statistics from the crystallographic analysis

I4 ₁ 22 form			P4 ₃ 2 ₁ 2 form		
Data set	Native	Thimerosal	Se-Met	Native	
Resolution (Å)	2.8	3.0	3.0	3.5	
Observations	169036	189542	114334	173168	
Unique reflections	35565	29905	28678	26690	
Data coverage (%)	95.4	98.3	94.1	69.1	
R _{merge} (%)	5.7	11.9	7.7	19.8	
R _{iso} (%)	–	13.8	9.8	–	
MIR phasing (15.0–3.0 Å)					
Phasing power					
Centric	–	0.40	0.65	–	
Acentric	–	0.55	0.91	–	
Mean FOM		0.299		–	
Refinement statistics					
	Resolution (Å)	Reflections	No of 2:2 complexes in asymmetric unit	Non-hydrogen atoms	
I4 ₁ 22 form	25.0–2.8	35,565	1	6,084	
P4 ₃ 2 ₁ 2 form	8.0–3.5	24,595	2	11,442	
R.m.s. deviations					
	R factor (%)	R _{free} (%)	Average B factor (Å ²)	Bonds (Å)	Angles (°)
I4 ₁ 22 form	22.1	29.8	54.2 (54.9)	0.014	1.9
P4 ₃ 2 ₁ 2 form	31.7	–	–	–	–
					B factor (Å ²)
I4 ₁ 22 form					7.2 (11.8)
P4 ₃ 2 ₁ 2 form					–

$R_{\text{merge}} = 100 \times \sum_i \sum_j |I_{ij} - \bar{I}_i| / \sum_i \sum_j I_{ij}$, where I_{ij} is the mean intensity of i th unique reflection with an index hkl , and $\bar{I}_i$ is the intensity of the i th observation of this reflection. $R_{\text{iso}} = 100 \times \sum_i |F(P_i) - F(P_i)| / \sum_i F(P_i)$, where $F(P)$ and $F(PH)$ are observed amplitudes of native and derivative structure factors, respectively. Phasing power = $\langle F(H) \rangle / E$, where $\langle F(H) \rangle$ is the r.m.s. amplitude of heavy-atom structure factor and E is the residual lack of closure error. Mean FOM = $\langle F(\text{best}) \rangle / \langle F(P_i) \rangle$, where $F(\text{best})$ is the amplitude of the best structure factor estimated by the MIR method. R -factor = $100 \times \sum_i |F(P_i) - F(\text{calc})| / \sum_i F(P_i)$, where $F(\text{calc})$ is the amplitude of a calculated structure factor based on the model. $R_{\text{free}} = R$ factor calculated using 5% of total reflections, which were chosen randomly and omitted from the refinement. All observed reflections, except for those for the calculation of R free, were used for the refinement. Average B factors were calculated from the refined model and the experimental Wilson plot (in parenthesis), respectively. R.m.s. deviations are deviations from ideal geometry. R.m.s. B factors were calculated from main-chain atoms and side-chain atoms (in parenthesis), respectively. Residues in disordered regions were eliminated from the model: T1–L3 (MOL1) and T1–P5 (MOL2) of GCSF; G214–H220 (MOL1) and K308–A309 (MOL1) of gs-CRH; A95 (MOL2); I213–H220 (MOL2), and C308–A309 (MOL2) of gs-CRH. Seven residues in GCSF (Q70 and L130 in MOL1; H52, W58, Q67, K69 and Q70 in MOL2) and six residues in gs-CRH (K157 and R158 in MOL1; K156, K157, R158 and Q221 in MOL2) were included in the model as alanine.

where the BC domain contacts the N-terminal region of the GCSF of the other 1:1 complex, connects the two 1:1 complexes (Fig. 1b, c). The absence of any other direct contacts between the receptors and the ligands in the 2:2 complex results in the formation of a solvent channel, occupied by an extensive hydrogen-bonded network of water molecules, at the centre of the complex with dimensions of about 15 Å high, 10 Å wide and 15 Å deep (Fig. 1b, c).

Although the individual components are essentially identical between the two 1:1 complexes (referred to as MOL1 and MOL2) in the asymmetric unit, the intermolecular orientations of the ligand and the receptor and the interdomain orientations within the receptor are substantially different (Fig. 1f). The difference in the BN–BC interdomain orientation between MOL1 and MOL2 is 3°, comparable to that reported for ligand-free gp130 (ref. 14). Even more striking, the orientation relative to GCSF differs by approximately 9° for the entire gs-CRH domain and by 10° and 8° for the BN and BC domains, respectively. The finding in another crystal form (Table 1) of relative orientations varying by only 2° implies that each 1:1 complex conformer is azimuthally restrained.

The major interface of the complex is characterized by extensive interactions between helices A and C of GCSF and a number of loops around the elbow region of gs-CRH (Fig. 1d), in agreement

with results from the alanine-scanning mutagenesis of the ligand¹⁶ and the receptor¹⁷. This interface corresponds to the weaker of the two recognition sites in the GH–GHR complex^{18,19}.

In the major interface of the MOL1 complex, ten polar interactions are made between six residues of GCSF and seven residues of gs-CRH, and 28 van der Waals contacts (within 4.2 Å) occur between 15 residues of GCSF and 15 residues of gs-CRH, with a buried surface area of 836 Å², whereas the MOL2 major interface has 14% fewer polar and van der Waals interactions but a comparable 801 Å² buried surface area. The excellent shape complementarity of the major interface, with the long side chains of residues such as Tyr, Arg, Lys, Glu and Gln stacking with each other to maximize the van der Waals contacts, is also found in the GH–GHR¹⁰ and EPO–EPOR¹¹ complexes, implying that such interactions are required for the recognition of long-chain class 1 cytokines by their receptors.

The core region of the interface is characterized by an interlaced network of hydrogen bonds which directly link amino acids, such as Glu 19 in GCSF to both Tyr 172 (BN) and Arg287 (BC) (Fig. 2), which have been shown to be critical for biological activity^{16,17}. Notably, the water molecules embedded in the major interface are not predominantly involved in mediating indirect hydrogen bonds but instead interact with residues that are already directly hydrogen

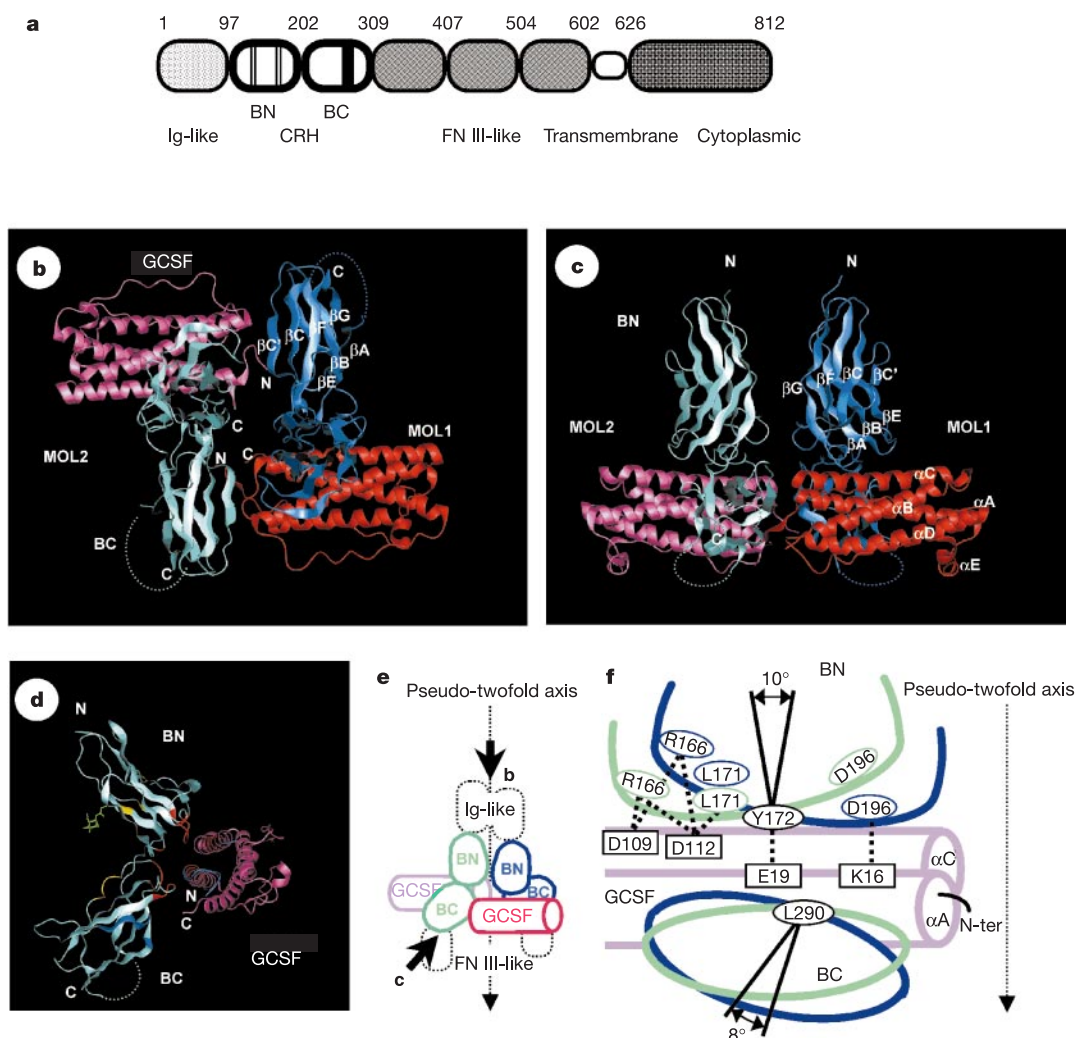


Figure 1 The GCSF–gs-CRH complex. MOL1 and MOL2 are shown in dark and light colours, respectively. **a**, Schematic drawing of the domain arrangements of GCSFR. The four highly conserved cysteines forming two disulphide bridges and the WSXWS motif in gs-CRH are indicated by double thin vertical and single thick vertical lines, respectively. **b**, **c**, Orthogonal views of the 2:2 complex. **d**, One 1:1 complex (MOL2) viewed from the right side of **b** and rotated 45° clockwise. The major (red) and minor contact (dark blue)

regions, the WSXWS motif (yellow), the four conserved cysteines (yellow) and the sugar moiety (green) are depicted. **e**, Drawing of the GCSF–gs-CRH complex with the **b** and **c** perspectives denoted. **f**, Drawing of the major interfaces of MOL1 and MOL2, showing the principal hydrogen-bonding linkages between GCSF and the BN domain of gs-CRH. Ig-like, immunoglobulin-like; N-ter, N terminus.

bonded across the interface, and therefore may contribute to reducing the thermodynamic penalty of desolvation in complex formation²⁰. The relatively even distribution of hydrophobic and polar interactions, which are also found in the IL-4–IL-4R complex¹⁵, indicates that polar interactions play more critical roles in assembly and subunit orientation than in the GH–GHR complex, where the core of the interface is hydrophobic with polar interactions at the perimeter.

The rotational differences between MOL1 and MOL2 are manifested through changes in the major interface. In particular, the BN and BC domains appear to pivot on the side chains of Tyr 172 and Leu 290, respectively (Fig. 1f). In addition, the residues, which are involved in interactions unique to one of the conformers, are localized on the BN side of the interface and are conserved, implying that the interactions in the BN domain are the dominant factor in determining the relative orientation. Interestingly, the destabilization of the MOL1 interface by alanine substitutions of certain residues involved in MOL1-specific interactions (Thr 115 and Glu 123 in GCSF, and Gln 173 and Asp 196 in gs-CRH) inexplicably caused a significant increase in the biological activity^{16,17}.

The minor interface is the pivotal factor in assembling the 1:1 complex into the 2:2 complex (Fig. 1b, c). With a buried surface area of 348 Å², an intermolecular anti-parallel β structure is formed through two main-chain hydrogen bonds between Phe 259 in the BC domain and Ser 7 in GCSF; as the N-terminal residues of GCSF alone are disordered¹², an induced fit conformational change upon ligand–receptor binding is inferred. In addition to a hydrogen bond between the main chain of Gln 11 in GCSF and the His 260 side chain in the BC domain and roughly ten van der Waals interactions, the network of ordered water molecules within the central solvent channel may contribute to the stability of the minor contact, as observed in antigen–antibody interfaces^{21,22}.

The legitimacy of the 2:2 complex is supported by the following evidence for the minor interface. First, Leu 261 in the minor interface showed a large chemical shift perturbation in an NMR titration study of the BC domain with GCSF¹³, and the relative ¹H–¹⁵N crosspeak intensities of the neighbouring backbone atoms decreased to the same extent as those in the major interface, indicating that the minor interface bound in the crystal structure does indeed exist in solution. Secondly, although all the deletion mutants of the GCSF N terminus cannot be explained, the reduction in biological activity after the deletion of the N-terminal segment up to Gln 11 (refs 23,

24) agrees with the involvement of residues from Ser 8 to Gln 11 in the minor interface. Finally, the robustness of the minor interface is indicated from the conservation of the interactions observed among the three different 2:2 complexes (Table 1).

The backbones of the BN and BC domains in gs-CRH are individually similar to those of the ligand-free gp130 (ref. 14); the r.m.s. deviations for the corresponding C α atoms in each domain are 1.0 Å for the BN domain and 2.5 Å for the BC domain. The larger value for the BC domain is consistent with its lower sequence homology. On the other hand, the interdomain orientations differ by 24–27° between the two receptors (Fig. 3), resulting apparently from differences in the loop containing the highly conserved WSXWS motif⁴ (the WSXWS loop). In gp130, but not gs-CRH, the BN–BC domains are tightly restrained by the main-chain hydrogen bonds between the WSXWS loop and the L strand (Fig. 3). Furthermore, the adoption in gs-CRH of the same interdomain orientation as that of gp130 would cause a serious collision at Leu 290 and Pro 291, which replace Gly 284 and Lys 285, respectively, in gp130. Notably, these two residues are conserved in the two receptor subgroups, respectively. Thus, the architecture of the WSXWS loop appears to be important for determining the interdomain orientation of the CRH region.

The present structure shows two additional crystallographic contacts between the BN and BC domains of separate 2:2 complexes and between the two GCSF molecules, although their overall features are not as strong as those in the major and minor interfaces. These contacts may participate in the association of the ligand and CRH region with the immunoglobulin-like domain. A plausible model in which the immunoglobulin-like domains simply extend from the BN domain and self-associate to stabilize the 2:2 complexes would not explain a discrepancy between our structure and a GCSF alanine scan¹⁶, which identified two potential recognition surfaces; although site 2 corresponds to the major or minor interface, all residues in site 1 are completely free from contact within the present 2:2 complex.

Several crystallographic studies of the EPOR domain complex¹¹, the EPO mimetic peptides–receptor domain complex²⁵ and the EPOR extracellular domain in the unliganded form²⁶, in combination with an *in vivo* functional assay²⁷, have highlighted the importance of the precise structural features, particularly the extracellular domain orientations which affect the separation between intracellular domains, and thus the inability of oligomeri-

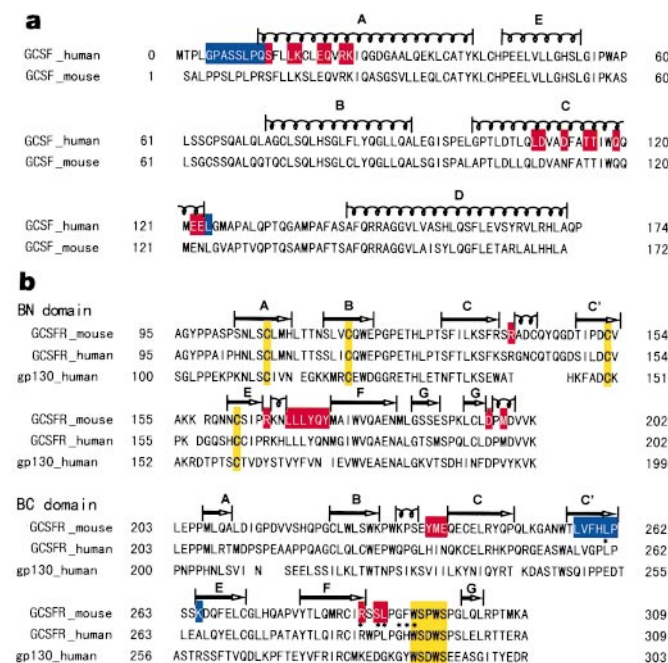


Figure 2 Secondary structures and sequences of GCSF and gs-CRH. Secondary structure assignment (PROCHECK²⁹) and multiple sequence alignments of human GCSF (a) and mouse gs-CRH (b). Helices and strands are indicated by coils and arrows, respectively. Red and blue shading denote the residues in the major and minor interfaces, respectively. Asterisks, residues in the BC domain detected in the NMR titration experiment¹³.

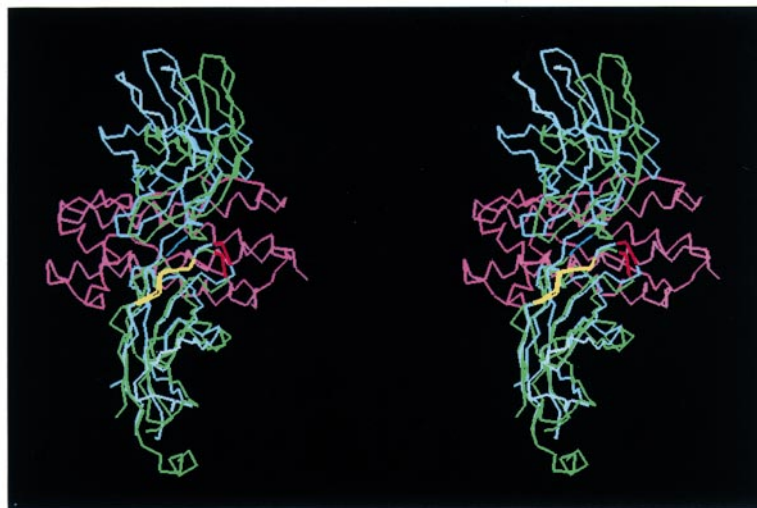


Figure 3 Structural comparison of the GCSF-gs-CRH complex and ligand-free gp130. Stereoview of the MOL2 GCSF-gs-CRH complex (pink and light blue) and gp130 (green) after an alignment of the corresponding C α atoms in the C-terminal domains of the

receptors. The residues corresponding to the WSXWS motif and to Leu 290-Pro 291 in MOL2 are coloured yellow and red, respectively. The L strand in gp130 is coloured dark blue.

zation of the receptor molecules to simply account for cellular activation. Therefore, the two distinct orientations in the current GCSF-gs-CRH complex may also be important in the formation of the fully active complex with the immunoglobulin-like domains from the partially active 2:2 complex. Thus, we believe that the current 2:2 complex crystal structure reflects one stage in the sequential formation of the active complex and should provide an important molecular basis for rational drug design. □

Methods

Data collection and heavy atom derivatives

Purification of mouse gs-CRH and crystallization of the complex with human GCSF will be described elsewhere (manuscript in preparation). Exchanging between the human and mouse sequences yields essentially no difference in either the K_d values or in the biological activities of the mixed ligand-receptor complex⁴. The crystals used for structural analysis belonged to the space group $I4_122$, with unit cell parameters of $a = b = 125.5$ Å, $c = 372.8$ Å. Derivatives were obtained by soaking crystals in 1 mM thimerosal for 6 h or by expression of GCSF in medium containing selenomethionine. For data collection, the cryosolvent was prepared by dissolving sucrose to 50% (w/v) in a harvest buffer (100 mM HEPES-NaOH and 1.2 M ammonium sulphate, pH 7.5). Transfer of the $I4_122$ crystal into a buffer of higher salt concentration (100 mM HEPES-NaOH, 2.0 M ammonium sulphate and 30% (v/v) glycerol, pH 7.5) induced conversion into a different crystal form ($P4_32_12$; $a = b = 125.7$ Å, $c = 373.3$ Å). The crystals were mounted in a nylon fibre loop and were flash-cooled in a nitrogen gas stream at 100 K. All diffraction data of the $I4_122$ crystals were collected at BL41XU in SPring-8 (Hyogo, Japan), with R-axis IV as the detector. The $P4_32_12$ crystal data were obtained at BL-6B in Photon Factor (Tsukuba, Japan). All of the data were processed by DENZO/SCALEPACK²⁸.

Phasing and refinement

The first mercury site was found in a difference Patterson map from the $I4_122$ crystal by RSPS²⁹. The other three mercury sites and six selenium sites were found in cross difference Fourier maps. Subsequently, the heavy atom parameters were refined by MLPHARE²⁹. Although the initial multiple isomorphous replacement (MIR) map was insufficient for chain tracing, the relative positions of the major mercury site and three of the selenium sites were consistent with the single free cysteine and the three methionine residues in the GCSF molecule¹². Application of a tentative overall NCS operator in DM²⁹ allowed us to build a model of the α -helical regions of GCSF and a part of the β -stranded regions of gs-CRH with the program QUANTA97 (Molecular Simulations). From this initial model, three NCS operators corresponding to each region (GCSF, BN and BC domains) were individually determined in order to apply multidomain molecular averaging for further phase improvement. The resulting map was of sufficient quality to build a model for the entire complex, including the side chains of gs-CRH. The model was refined with X-PLOR³⁰ and REFMAC²⁹ programs, and with subsequent manual rebuilding. Bulk-solvent correction, overall anisotropic scaling and NCS restraints were applied. The current model contains 750 amino-acid residues, 260 water molecules and two *N*-acetyl-glucosamine residues attached to the two NCS-related Asn 104s of gs-CRH. Asn 104 was the only glycosylation site found by chemical analysis (manuscript in preparation).

The molecular replacement analysis in AMORE²⁹ allowed the determination of the $P4_32_12$ crystal containing the two 2:2 complex molecules, which are arranged in a slightly different manner from those in the $I4_122$ crystal. The structure was subjected to the rigid-body refinement of X-PLOR. A stereochemical analysis of the $I4_122$ form by PROCHECK²⁹

showed that 84.4% of the non-glycine and non-proline residues are lying in the most favoured regions of the Ramachandran-plot, while 1.3% residues are in the generously allowed region. Only two residues (NCS-related Ile 95s of GCSF) in the disallowed areas showed essentially the same ϕ/ψ angles in the GCSF structure determined at 2.2 Å resolution¹². The buried surface area was calculated in QUANTA97 with a probe radius of 1.4 Å; the area was defined as a half of an accessible surface area, obtained by subtracting that of the complex from the summed area of the two components. MOL1 and MOL2 were oriented by superimposing the C α atoms of GCSF (Fig. 1f), and then the rotational angle χ of the spherical polar angle was derived from the best superposition of the two components by the program LSQKAB²⁹. Figures 1 and 3 were produced with the program QUANTA97.

Received 20 May; accepted 6 August 1999.

1. Metcalf, D. The molecular control of cell division, differentiation commitment and maturation in haemopoietic cells. *Nature* **339**, 27–30 (1989).
2. Welte, K. *et al.* Filgrastim (r-metHuG-CSF): the first 10 years. *Blood* **88**, 1907–1929 (1996).
3. Wells, J. A. & de Vos, A. M. Hematopoietic receptor complexes. *Annu. Rev. Biochem.* **65**, 609–634 (1996).
4. Fukunaga, R., Ishizaka-Ikeda, E., Pan, C.-X., Seto, Y. & Nagata, S. Functional domains of the granulocyte colony-stimulating factor receptor. *EMBO J.* **10**, 2855–2865 (1991).
5. Hiraoka, O., Anaguchi, H. & Ota, Y. Formation of 1:1 complex of the cytokine receptor homologous region of granulocyte colony-stimulating factor receptor with ligand. *Biosci. Biotechnol. Biochem.* **59**, 2351–2354 (1995).
6. Hiraoka, O., Anaguchi, H., Asakura, A. & Ota, Y. Requirement for the immunoglobulin-like domain of granulocyte colony-stimulating factor receptor in formation of a 2:1 receptor–ligand complex. *J. Biol. Chem.* **270**, 25928–25934 (1995).
7. Horan, T. P. *et al.* Coexpression of G-CSF with an unglycosylated G-CSF receptor mutant results in secretion of a stable complex. *Protein Express. Purific.* **14**, 45–53 (1998).
8. Hiraoka, O., Anaguchi, H. & Ota, Y. Evidence for the ligand-induced conversion from a dimer to a tetramer of the granulocyte colony-stimulating factor receptor. *FEBS Lett.* **356**, 255–260 (1994).
9. Horan, T. P. *et al.* Dimerization of the extracellular domain of granulocyte-colony stimulating factor receptor by ligand binding: a monovalent ligand induces 2:2 complexes. *Biochemistry* **35**, 4886–4896 (1996).
10. de Vos, A. M., Ultsch, M. & Kossiakoff, A. A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312 (1992).
11. Syed, R. S. *et al.* Efficiency of signaling through cytokine receptors depends critically on receptor orientation. *Nature* **395**, 511–516 (1998).
12. Hill, C. P., Osslund, T. D. & Eisenberg, D. The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. *Proc. Natl Acad. Sci. USA* **90**, 5167–5171 (1993).
13. Yamasaki, K., Naito, S., Anaguchi, H., Ohkubo, T. & Ota, Y. Solution structure of an extracellular domain containing the WSXWS motif of the granulocyte colony-stimulating factor receptor and its interaction with ligand. *Nature Struct. Biol.* **4**, 498–504 (1997).
14. Bravo, J., Staunton, D., Heath, J. K. & Jones, E. Y. Crystal structure of a cytokine-binding region of gp130. *EMBO J.* **17**, 1665–1674 (1998).
15. Hage, T., Sebald, W. & Reinemer, P. Crystal structure of the Interleukin-4/Receptor α chain complex reveals a mosaic binding interface. *Cell* **97**, 271–281 (1999).
16. Reidhaar-Olson, J. F., De Souza-Hart, J. A. & Selick, H. E. Identification of residues critical to the activity of human granulocyte colony-stimulating factor. *Biochemistry* **35**, 9034–9041 (1996).
17. Layton, J. E., Iaria, J., Smith, D. K. & Treutlein, H. R. Identification of a ligand-binding site on the granulocyte colony-stimulating factor receptor by molecular modeling and mutagenesis. *J. Biol. Chem.* **272**, 29735–29741 (1997).
18. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone–receptor interface. *Science* **267**, 383–386 (1995).
19. Wells, J. A. Binding in the growth hormone receptor complex. *Proc. Natl Acad. Sci. USA* **93**, 1–6 (1996).

20. Hendsch, Z. S. & Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226 (1994).
21. Braden, B. C. & Poljak, R. J. Structural features of the reactions between antibodies and protein antigens. *FASEB J.* **9**, 9–16 (1995).
22. Covell, D. G. & Wallqvist, A. Analysis of protein–protein interactions and the effects of amino acid mutations on their energetics: the importance of water molecules in the binding epitope. *J. Mol. Biol.* **269**, 281–297 (1997).
23. Kuga, T. *et al.* Japanese Patent Publication (Kokoku) 8317 Feb. 2 (1994).
24. Yamasaki, M., Konishi, N., Yamaguchi, K., Itoh, S. & Yokoo, Y. Purification and characterization of recombinant human granulocyte colony-stimulating factor (rhG-CSF) derivatives: KW-2228 and other derivatives. *Biosci. Biotechnol. Biochem.* **62**, 1528–1534 (1998).
25. Livnah, O. *et al.* An antagonist peptide–EPO receptor complex suggests that receptor dimerization is not sufficient for activation. *Nature Struct. Biol.* **5**, 993–1004 (1998).
26. Livnah, O. *et al.* Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* **283**, 987–990 (1999).
27. Remy, I., Wilson, I. A. & Michnick, S. W. Erythropoietin receptor activation by a ligand-induced conformation change. *Science* **283**, 990–993 (1999).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
30. Brünger, A. T. X-PLOR, Version 3.1, *A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992).

Acknowledgements

We thank N. Kamiya and M. Kawamoto for their help in the use of the data collection facilities at SPring-8, Hyogo, Japan, and N. Sakabe and N. Watanabe for use of the facilities at Photon Factory, Tsukuba, Japan. We thank S. Nagata for providing the GCSFR cDNA, E. Yanagihara and N. Okitsu for assistance in sample preparation, and Y. Kato, D. Kohda, H. Toh, K. Yamasaki and S. Tsutakawa for useful discussions. This study was partly supported by a grant from TARA (Tsukuba Advanced Research Alliance), University of Tsukuba, Japan.

Correspondence and requests for materials should be addressed to K.M. (e-mail: morikawa@beri.co.jp). Coordinates have been deposited in the Brookhaven Protein Data Bank, accession codes 1cd9 for the I4₂₂ and 1pgr for the P4₃2₂ form.

Structural evidence for dimerization-regulated activation of an integral membrane phospholipase

H. J. Snijder*, I. Ubarretxena-Belandia†‡, M. Blaauw*‡, K. H. Kalk*, H. M. Verheij†§, M. R. Egmond†, N. Dekker† & B. W. Dijkstra*

* Laboratory of Biophysical Chemistry, BIOSON Research Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands
 † Department of Enzymology and Protein Engineering, Centre for Biomembranes and Lipid Enzymology, Institute of Biomembranes, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Dimerization is a biological regulatory mechanism employed by both soluble and membrane proteins¹. However, there are few structural data on the factors that govern dimerization of membrane proteins. Outer membrane phospholipase A (OMPLA) is an integral membrane enzyme which participates in secretion of colicins in *Escherichia coli*. In *Campylobacter*² and *Helicobacter pylori* strains³, OMPLA is implied in virulence. Its activity is regulated by reversible dimerization^{4,5}. Here we report X-ray structures of monomeric and dimeric OMPLA from *E. coli*. Dimer interactions occur almost exclusively in the apolar membrane-embedded parts, with two hydrogen bonds within the hydrophobic membrane area being key interactions. Dimerization results in functional oxyanion holes and substrate-binding pockets, which are absent in monomeric OMPLA. These results

provide a detailed view of activation by dimerization of a membrane protein.

The structure of monomeric OMPLA consists of a 12-stranded antiparallel β -barrel with a convex and a flat side (Fig. 1a). The 12 amphipathic β -strands traverse the membrane, forming a hydrophobic outer surface flanked by rings of aromatic residues at the polar–apolar boundaries of the membrane bilayer. Loops (L1–L6) join consecutive strands at the extracellular side, and turns (T1–T5) connect the strands at the periplasmic side of the membrane (as inferred from epitope insertion studies⁶). The interior of the β -barrel is polar and contains an intricate hydrogen-bonding network providing a rigid structure. The two termini and the L1, L4 and L6 loops cover the barrel's interior and obstruct pore function, as corroborated by the absence of pore activity in black-lipid films (N.D., unpublished results). The 12- β -stranded architecture of OMPLA resembles the folds of other β -barrel outer membrane proteins^{7–11} that have 8–22 β -strands. This shows the functionality of this fold not only for transport and structural proteins but also for membrane enzymes.

The active site contains Ser 144 and His 142, as established from chemical modification¹² and site-directed mutagenesis experiments^{13,14}, and is located on the exterior of the β -barrel, positioned just outside the outer leaflet ring of aromatic residues. Besides the serine and histidine, the active site contains an asparagine residue (Asn 156). Its side chain is at hydrogen-bonding distance from the N δ 1 atom of His 142, fixing the orientation of the His 142 side chain. This constellation of active-site residues resembles that of classical serine hydrolases¹⁵, but with a neutral asparagine instead of an aspartate residue. Mutation studies have confirmed the importance of Asp 156 for catalysis (R. L. Kingma *et al.*, manuscript in preparation).

Reversible dimerization regulates the activity of OMPLA, with the dimer being the active species^{4,5}. Covalent inhibition with hexadecanesulphonylfluoride allowed the crystallization of the dimeric enzyme. The structural rearrangements upon dimerization are surprisingly few and small (r.m.s. difference between monomer and dimer is 0.4 Å for 257 C α atoms). The monomers (labelled A and B) associate with the flat sides of their β -barrels to form a homodimer (Fig. 2a), such that both active sites are located at the outer leaflet side of the membrane. Roughly 1,100 Å² of the accessible surface is buried, which corresponds to 10% of the total surface. All interactions between the monomers occur in the hydrophobic membrane-embedded area (Fig. 2b), except for three main-chain hydrogen bonds formed at the polar–apolar membrane boundary (Phe 109^A–Gly 146^B, Gly 146^A–Phe 109^B and Leu 32^A–Leu 32^B). The hydrophobic side chains of Leu 32^{A/B}, Leu 71^{A/B}, Leu 73^{A/B} and Leu 265^{A/B} exhibit a knob-and-hole interaction. Residues Tyr 114^{A/B}–Phe 109^{B/A} have aromatic stacking interactions with contact distances as close as 3.5 Å. The tyrosine hydroxyl groups point towards a cavity with a hydrophilic character extending deep into the hydrophobic membrane-embedded area. The cavity is lined with the side chains of Tyr 92^{A/B}, Gln 94^{A/B}, Ser 96^{A/B} and Thr 112^{A/B}. In monomeric OMPLA, a stabilizing hydrogen-bond network between Tyr 92, Gln 94 and Ser 96 allows these residues to partition in the hydrophobic membrane environment. In the dimer, however, the side chain of Gln 94 is additionally hydrogen bonded to Gln 94 of the other monomer (Fig. 1b). Gln 94 is strictly conserved in all OMPLA sequences, emphasizing the importance of this residue for functional dimerization. Thus, specific dimer interactions are obtained from a combination of surface complementarity of knob-and-hole patterns and hydrogen-bonding interactions between polar residues in the hydrophobic membrane environment. Similar polar patches have been observed in the apolar membrane-embedded regions of other membrane proteins^{8,10,11}, but here we show for the first time, to our knowledge, their participation in protein–protein interactions.

Dimerization results in productive active sites, with two substrate-binding pockets created at the interface between the two

‡ Present addresses: Molecular Biophysics and Biochemistry, Yale University, 429 Bass Center, 266 Whitney Avenue, New Haven, Connecticut 06520, USA (I.U.B.); Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands (M.B.).

§ H.M.V. died in an accident on 1 August 1998.

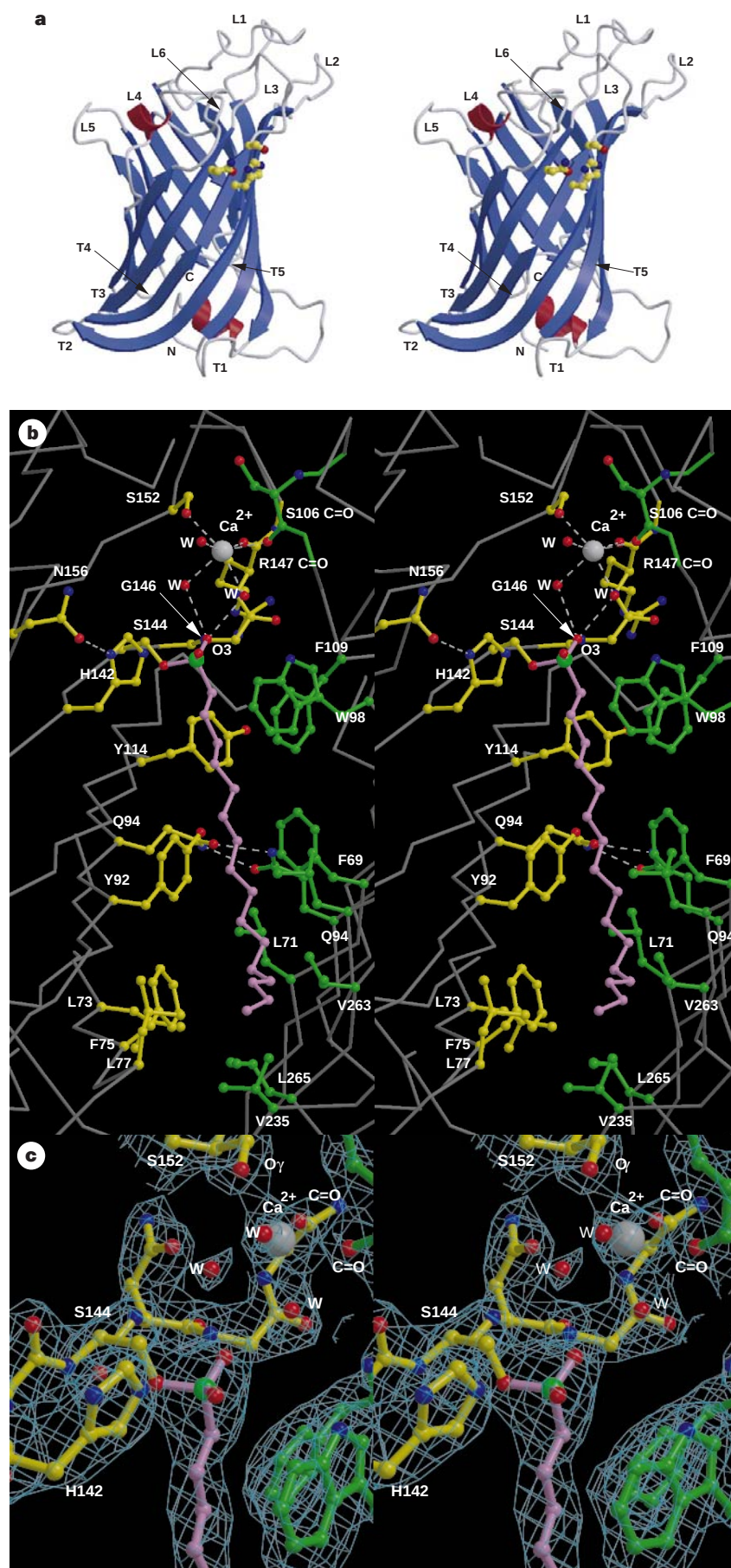


Figure 1 Structure of outer membrane phospholipase A. **a**, Stereo drawing of the secondary structure of a monomer. The α -helices, β -strands and turns are depicted in red, blue and grey, respectively. Loops and turns are numbered from the N to C terminus and defined as L1, residues 46–64; L2, 99–108; L3, 145–153; L4, 179–196; L5, 215–221; L6, 246–254; T1, 80–85; T2, 129–130; T3, 167–168; T4, 204–205; and T5, 231–235. The active-site residues (ball-and-stick representation) are located at the exterior of the β -barrel, facing the outer leaflet side of the membrane. **b**, Stereo diagram of the active site and binding pocket of OMPLA. The inhibitor and residues from monomer A and B are coloured in purple, yellow and green, respectively. Dashed lines indicate hydrogen bonds in the active site; in the calcium-binding site and between the two glutamines Gln 94^{A/B} involved in dimer interaction. The calcium ion is shown; unexpectedly, none of its ligands is charged. **c**, Stereo diagram of the electron density in the active site of dimeric OMPLA with $2F_o - F_c$ OMIT electron density contoured at 1σ level.

monomers. Both pockets accommodate all 16 aliphatic carbon atoms of the inhibitors (Figs 1b and 2b). Aromatic residues Tyr 114^A, Phe 109^B and Trp 98^B contact the upper three carbon atoms of the inhibitor. The side chains of Tyr 92^A, Phe 69^B and Leu 40^B line the substrate-binding site for carbon atoms six to eight of the inhibitor. After a bend at residues Leu 71^B and Leu 265^B (Figs 1b and 2b), the remainder of the binding pocket is lined with residues Phe 75^A, Leu 77^A, Val 235^B and Val 236^B, illustrating the exclusively hydrophobic character of the binding pocket. Both monomers contribute to binding the inhibitor. In particular, contacts for carbon atoms 10–16 are provided by the neighbouring molecule. Binding of the inhibitors accounts for 27% of the dimerization surface, which rationalizes the observed increase in dimer stability of the inhibited enzyme⁴.

The catalytic sites are located at the top end of each substrate-binding pocket, where a cup-like structure is formed at the interface of the monomers (Fig. 1b). The base of the active site consists of Tyr 114^A, Trp 98^B and Phe 109^B. The walls are formed by the active site His 142^A, the amino-terminal residues of β -strand 6^A and loops L2^B, L3^A and L4^A. The sulphonyl moiety of the inhibitor is located in the centre of this active-site cup, where it is covalently bound to the O γ of the nucleophilic serine (Fig. 1c). It mimics the transient tetrahedral intermediate of the deacylation step in the serine hydrolase catalytic mechanism. The O3 sulphonyl oxygen atom is at hydrogen-bonding distance from the main-chain nitrogen atom of Gly 146^A and two^A (or one^B) water molecules. These water molecules are bound, in turn, to a calcium ion present in the active site. Calcium is an essential cofactor for activity¹⁶ and dimerization⁴. Its binding site is located at the dimer interface and the ion has ligands from both monomers (main-chain carbonyl oxygen atoms from Arg 147^A and Ser 106^B, the O γ atom of Ser 152^A and three water molecules; Figs 1c and 3). Ser 152 is essential for catalysis¹⁴. The structure indicates that the catalytic importance of this residue may be related to its calcium-ligating capacity. Despite biochemical evidence for a second calcium site¹⁶, we find only one bound calcium ion per monomer, presumably caused by the enzyme's low calcium affinity in the presence of uncharged detergents¹⁶.

A phospholipid substrate may bind to OMPLA in a similar fashion to the inhibitor, with one acyl chain occupying the position of the inhibitor's aliphatic chain. The second acyl chain may be located in the membrane compartment without the need for a particular binding site. The phosphate group of the substrate

probably extends towards the solvent and could bind to the calcium ion; however, catalysis requires neither the second acyl chain nor the phosphate group¹⁷, explaining the absence of obvious binding sites for them. Hydrolysis of the substrate could proceed through a mechanism analogous to that of serine hydrolases, with Ser 144 performing a nucleophilic attack on the carbonyl carbon of the ester. A negatively charged intermediate will then be formed which is stabilized by a tetrahedral arrangement of hydrogen bonds (the amide of Gly 146 and two water molecules) and by the remote influence of the calcium ion (at 4.3 Å). The transient intermediate will then collapse to give the enzyme-acyl intermediate. The lyso-phospholipid product can now leave the active site by lateral diffusion into the membrane. Likewise, deacylation occurs with a water molecule acting as nucleophile. The fatty-acid product may leave the active site either by lateral diffusion into the membrane through the opening between His 142^A and Trp 98^B (Fig. 1b) or after dissociation of the dimer. Calcium may have a triple role in catalysis. It may increase the nucleophilicity of the serine, it could polarize the ester carbonyl bond, thereby facilitating nucleophilic attack, and it is involved in oxyanion stabilization. A comparable active site and well-developed oxyanion stabilization have been observed in *Streptomyces scabies* esterase¹⁸, which relies on a His-Ser dyad with an carbonyl oxygen atom stabilizing the histidine, and three tetrahedrally arranged hydrogen bonds coordinating the oxyanion (Fig. 3).

Both calcium⁴ and substrate¹⁹ govern dimerization. The main factor that controls dimerization *in vivo* is probably the substrate. Normally, phospholipids are only present in the inner leaflet of the outer membrane, and OMPLA is in the inactive monomeric form⁵. As activation of OMPLA concurs with the perturbation of the bacterial membrane⁴, we propose that the relief of lipid asymmetry and the presentation of substrate to the outer leaflet induces activation by promoting dimerization and calcium binding. The OMPLA dimer then hydrolyses the substrate into lyso-phospholipids and fatty acids. These compounds will further destabilize the membrane bilayer, and may thus facilitate the release of colicins or virulence factors.

In conclusion, our work shows that, even in the membrane-embedded parts of integral membrane proteins, polar interactions can contribute to specific intermolecular contacts. These can be seen as the membrane equivalent of burying hydrophobic side chains in the binding interface of soluble proteins. Our findings could aid in the understanding of a wide variety of activation and regulation

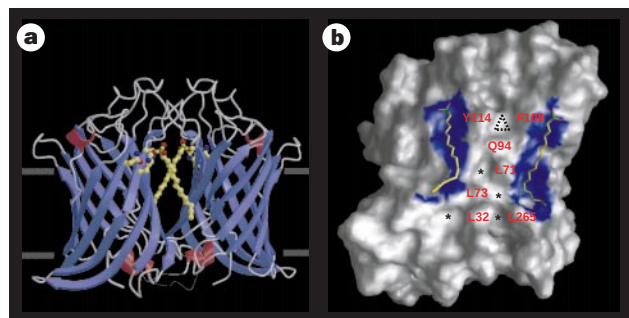


Figure 2 Dimer covalently inhibited with hexadecanesulphonylfluoride. **a**, Diagram of dimer with active-site residues and inhibitor shown in ball-and-stick representation. Residues 25–31 which lack interpretable density are indicated by thin dashed lines. The polar-apolar membrane boundaries are roughly indicated in grey. **b**, Surface representation of the dimerization interface of one monomer. The surface area within 1.5 Å of the van der Waals surface of the inhibitors is depicted in blue. The inhibitor on the left is covalently bound to Ser 144 of the monomer shown; the inhibitor on the right belongs to the monomer that is omitted in this diagram. Residues that are involved in the dimer interaction in the membrane-embedded part are indicated. Asterisks indicate the holes from the knob-and-hole pattern. The dashed triangle indicates the hydrophilic cavity (see text).

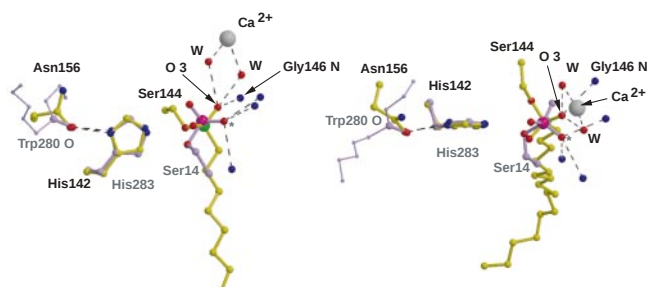


Figure 3 Comparison of the catalytic machinery of OMPLA and *Streptomyces scabies* esterase. Two orthogonal views show superimposition of the imidazole rings and carboxamide (Asn 156) and carbonyl (Trp 283) oxygen atoms; however, the orientation of the serine residues, inhibitors and oxyanion holes is mirrored with respect to each other. OMPLA is shown in yellow with black labels; *S. Scabies* esterase is shown in purple with grey labels. Oxyanion-hole ligands are shown but labelled only for OMPLA; the oxygen O3 atom proposed to correspond to the oxyanion is indicated. The corresponding oxygen in the *S. scabies* esterase is indicated with an asterisk.

Table 1 Data collection and phasing statistics

	Monomer			Ir-3 λ1	Ir-3 λ2	Ir-3 λ3	Se-Met λ1	Se-Met λ2	Se-Met λ3	Dimer
	Set 1	Set 2	Set 3							
Resolution (Å)	36.6–2.5	18.1–2.2	3.5–2.2	20.0–3.2	20.0–3.1	20.0–30	19.0–2.8	19.0–2.8	19.0–2.8	20.3–2.1
Wavelength (Å)	0.9164	0.9901	0.9901	1.10394	1.10444	0.827	0.97891	0.97969	0.90	0.99385
No. of unique, reflections	10973	13581	7532	6585	7244	7598	9365	9013	8161	28336
Redundancy	8.4	5.9	7.0	7.3	7.1	5.9	3.4	3.7	2.9	2.69
Completeness (%)	81.7 (42.6)	69.4 (4.1)	51.9 (5.2)	97.1 (92.1)	93.2 (81.2)	93.2 (77.8)	95.5 (95.8)	96.8 (93.5)	87.5 (91.8)	74.1 (66.3)
R _{sym} * (%)	4.1 (33.6)	4.9 (15.9)	9.0 (13.4)	9.8 (26.3)	7.1 (26.1)	6.1 (29.9)	5.4 (28.3)	5.9 (28.5)	5.7 (31.8)	7.7 (26.5)
⟨I⟩/⟨s(I)⟩	17.2 (2.9)	18.1 (2.8)	10.9 (3.7)	15.7 (2.6)	22.0 (4.2)	25.0 (3.7)	14.7 (3.3)	15.6 (3.5)	11.5 (2.4)	10.9 (2.3)
R _{iso} † (%)				18.2	17.1	18.6	14.8	14.8	16.2	
No. of sites				2	2	2	5	5	5	
Phasing power‡ acentric/centric				1.95/1.12	1.96/1.50	2.13/1.60	1.23/0.77	1.15/0.90	1.51/1.00	
R _{cullis} § iso				0.67/0.74	0.66/0.75	0.64/0.72	0.81/0.77	0.83/0.72	0.77/0.81	
R _{cullis} § ano				0.79	0.80	0.78	0.78	0.93	0.87	
Beam line	X11	D2AM	D2AM	X31	X31	X31	X31	X31	X31	ID2B
Refinement statistics										
			Monomer							Dimer
Resolution range (Å)			36.6–2.2							20.3–2.1
Total no. of nonhydrogen atoms			2231							4220
No. of water molecules			28							69
R-factor (%)			22.1							22.6
Free R-factor (%)			27.0							28.3
Average B-factor (Å ²)			56.0							34.5
R.m.s.d. bond lengths (Å)			0.005							0.007
R.m.s.d. bond angles			1.3°							1.2°

* $R_{sym} = \sum_{hkl} \sum_{sym} |F - \langle F \rangle| / \sum_{hkl} F$.
† $R_{iso} = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} F_{calc}$.
‡ Phasing power = $\sum_{hkl} |F_h| / \sum_{hkl} |E| = \sum_{hkl} |F_h| / \sum_{hkl} |F_{ph_{calc}}| - |F_{ph_{obs}}| = \langle F_r \rangle / \langle \sigma \rangle$.
§ $R_{cullis} = \sum_{hkl} |F_{ph_{obs}}| - |F_{ph_{calc}}| / \sum_{hkl} |F_{ph_{obs}}| - |F_{ph_{calc}}|$ isomorphous = $\langle \sigma \rangle / \langle \sigma \rangle$ (diff).
§ $R_{cullis} = \sum_{hkl} |F^+ - F^-| - (2 \times |F_h^{calc}|) / \sum_{hkl} |F^+ - F^-|$ anomalous = $\langle \sigma \rangle / \langle \sigma \rangle$ (ano. diff).

processes, such as signal transduction, phospholipid metabolism and bacterial pathogenicity.

Methods

Crystallization

Crystals of monomeric OMPLA (relative molecular mass 31K; 269 amino acids) were obtained by mixing protein solution (8 mg ml⁻¹ OMPLA, 10 mM succinate buffer pH 6.5, 1.5% 1-*O*-*n*-octyl-β-D-glucopyranoside (β-OG), 1 mM Na₂S₂O₃) in a 3:2 ratio with a reservoir solution containing 25–29% (v/v) 2-methyl-2,4-pentanediol, 1 mM CaCl₂ and 0.1 M bis-Tris pH 5.9–6.0 (ref. 20). The space group of the crystals was P3₁21 with cell dimensions *a* = *b* = 78.55, *c* = 101.52 Å. The protein labelled with selenomethionine crystallized under similar conditions to the wild-type protein.

Crystals of dimeric OMPLA were obtained by covalent modification of the active-site serine by the inhibitor hexadecanesulphonylfluoride¹² before crystallization. Crystals were grown from 10 mg ml⁻¹ inhibited protein solution containing 1.95% (w/v) β-OG, 10 mM KCl, 2 mM Tris pH 6.6 equilibrated against 12–17% (v/v) PEG400. Crystals of the inhibitor–enzyme complex were soaked for 20 min in a solution containing 0.5% (w/w) β-OG, 70% (w/w) phytohistol (Roth) and 15 mM CaCl₂ before data collection. These crystals were flash frozen using phytohistol as a cryoprotectant. The crystals belong to space group P2₁2₁2₁ with cell dimensions *a* = 81.54 Å, *b* = 84.97 Å and *c* = 95.80 Å, and have two molecules in the asymmetric unit.

Data collection and processing

Native data of monomeric OMPLA were collected on cryopreserved crystals at EMBL beam line X11, DESY, Hamburg (data set 1) and at the D2AM beam line at the ESRF, Grenoble (data sets 2 and 3). Data were processed using DENZO/SCALEPACK²¹ for data set 1, and XDS²² for data sets 2 and 3. The three data sets were merged using KBRANI (BIOMOL software package); data collected on the D2AM beam line were merged first (*R*_{merge} 7.7%) and subsequently merged with data set 1 (*R*_{merge} 10.5%). The crystals diffracted anisotropically, resulting in an effective resolution of the merged data of about 2.4 Å. An iridium derivative was prepared by soaking a native crystal in stabilizing mother liquor at pH 6.5 with 10 mM Na₂IrCl₆ for one day, which gave two unique sites. Three-wavelength anomalous dispersion experiments were collected on the iridium and selenomethionine derivatives at EMBL beam line X31, DESY, Hamburg. These data were processed with DENZO/SCALEPACK²¹ and scaled with KBRANI. Data on a dimer crystal was collected to 2.1 Å at the ID2B beam line at the ESRF, Grenoble, and processed with DENZO/SCALEPACK²¹. Crystallographic data are summarized in Table 1.

Structure determination

Isomorphous differences between derivatives and native data were included in phasing, in addition to the MAD phase information. Phases calculated with MLPHARE²³ were improved by solvent flipping and histogram matching using DM²³. The resulting electron-density map allowed placement of residues 13–269. The model was refined using the slow-cooling protocol and conjugate gradient positional refinement in X-PLOR²⁴ cycled with manual rebuilding using O²⁵. Bulk solvent correction, overall anisotropic *B*-factor scaling

and a resolution-dependent weighting were applied, and temperature factors were refined individually. The model further comprises 28 water molecules and 5 partly occupied β-OG molecules. The final *R* and *R*_{free} of the model are 22.0% and 27.1%, respectively. Remaining difference density around the protein molecule was interpreted as disordered detergent. 84.8% of the main-chain phi–psi torsion angles is in the most favoured regions, whereas none is in disallowed regions of the Ramachandran plot.

Phases for the dimer were obtained by molecular replacement (AMoRE²³) using the monomer as a search model. The dimer was refined using a slow-cooling protocol and conjugate gradient positional refinement in X-PLOR²⁴ cycled with minor rebuilding using O²⁵. Bulk solvent correction, overall anisotropic *B*-factor scaling, and a resolution-dependent weighting were applied. Initially, tight non-crystallographic symmetry (NCS) restraints were applied to the entire molecules, but later these restraints were gradually relieved. Water molecules were included in the model. Two water molecules with exceptionally low *B*-factors and substantial difference density were interpreted as calcium ions. Restrained individual temperature factor refinement resulted in a final *R* of 22.6% and *R*_{free} of 28.3%. Molscript²⁶ was used to produce Figs 1a, b, 2a and 3; Raster3D²⁷ was used for all figures except Fig. 2b; GRASP²⁸ was used for Fig. 2b; and the program Bobscript²⁹ was used for Fig. 1c.

Received 7 May; accepted 6 August 1999.

1. Klemm, J. D., Schreiber, S. L. & Crabtree, G. R. Dimerization as a regulatory mechanism in signal transduction. *Annu. Rev. Immunol.* **16**, 569–592 (1998).
2. Grant, K. A., Ubarretxena-Belandia, I., Dekker, N., Richardson, P. T. & Park, S. F. Molecular characterization of *pldA*, the structural gene for a phospholipase A from *Campylobacter coli*, and its contribution to cell-associated hemolysis. *Infect. Immun.* **65**, 1172–1180 (1997).
3. Bukholm, G. *et al.* Colony variation of *Helicobacter pylori*: pathogenic potential is correlated to cell wall lipid composition. *Scan. J. Gastroenterol.* **32**, 445–454 (1997).
4. Dekker, N., Tommassen, J., Lustig, A., Rosenbusch, J. P. & Verheij, H. M. Dimerization regulates the enzymatic activity of *Escherichia coli* outer membrane phospholipase A. *J. Biol. Chem.* **272**, 3179–3184 (1997).
5. Dekker, N., Tommassen, J. & Verheij, H. M. Bacteriocin release protein triggers dimerization of outer membrane phospholipase A *in vivo*. *J. Bacteriol.* **181**, 3281–3283 (1999).
6. Merck, K. B., de Cock, H., Verheij, H. M. & Tommassen, J. Topology of the outer membrane phospholipase A of *Salmonella typhimurium*. *J. Bacteriol.* **179**, 3443–3450 (1997).
7. Weiss, M. S. *et al.* Molecular architecture and electrostatic properties of a bacterial porin. *Science* **254**, 1627–1630 (1991).
8. Schirmer, T., Keller, T. A., Wang, Y.-F. & Rosenbusch, J. P. Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science* **267**, 512–514 (1995).
9. Pautsch, A. & Schulz, G. E. Structure of the outer membrane protein A transmembrane domain. *Nature Struct. Biol.* **5**, 1013–1017 (1998).
10. Locher, K. P. *et al.* Transmembrane signaling across the ligand-gated FhuA receptor: Crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell* **95**, 771–778 (1998).
11. Buchanan, S. K. *et al.* Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. *Nature Struct. Biol.* **6**, 56–63 (1999).
12. Horrevorts, A. J., Verheij, H. M. & de Haas, G. H. Inactivation of *Escherichia coli* outer-membrane phospholipase A by the affinity label hexadecanesulfonyl fluoride. Evidence for an active-site serine. *Eur. J. Biochem.* **198**, 247–253 (1991).
13. Brok, R. G. P. M., Dekker, N., Gerrits, N., Verheij, H. M. & Tommassen, J. A conserved histidine

- residue of *Escherichia coli* outer-membrane phospholipase A is important for activity. *Eur. J. Biochem.* **234**, 934–938 (1995).
14. Brok, R. G. P. M., Ubarretxena-Belandia, I., Dekker, N., Tommassen, J. & Verheij, H. M. *Escherichia coli* outer membrane phospholipase A: Role of two serines in enzymatic activity. *Biochemistry* **35**, 7787–7793 (1996).
 15. Dodson, G. & Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347–352 (1998).
 16. Ubarretxena-Belandia, I., Boots, J. W. P., Verheij, H. M. & Dekker, N. Role of the cofactor calcium in the activation of outer membrane phospholipase A. *Biochemistry* **37**, 16011–16018 (1998).
 17. Horrevoets, A. J. G., Hackeng, T. M., Verheij, H. M., Dijkman, R. & De Haas, G. H. Kinetic characterization of *Escherichia coli* outer membrane phospholipase A using mixed detergent-lipid micelles. *Biochemistry* **28**, 1139–1147 (1989).
 18. Wei, Y. *et al.* A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nature Struct. Biol.* **2**, 218–223 (1995).
 19. Ubarretxena-Belandia, I. *et al.* Outer membrane phospholipase A is dimeric in phospholipid bilayers: A cross-linking and fluorescence resonance energy transfer study. *Biochemistry* **38**, 7398–7405 (1999).
 20. Blaauw, M., Dekker, N., Verheij, H. M., Kalk, K. H. & Dijkstra, B. W. Crystallization and preliminary X-ray analysis of outer membrane phospholipase A from *Escherichia coli*. *FEBS Lett.* **373**, 10–12 (1995).
 21. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
 22. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
 23. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
 24. Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of structures. *Nature* **355**, 472–475 (1992).
 25. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
 26. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
 27. Merritt, E. A. & Bacon, J. A. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
 28. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
 29. Esnouf, R. M. An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 132–134 (1997).

Acknowledgements

We thank the staff at the EMBL-outstation at DESY in Hamburg, and at the D2AM and the ID2B beam lines at ESRF Grenoble, for assistance with data collection. We thank R. L. Kingma for her enthusiastic collaboration and discussions. We thank the ESRF for supporting the work at the ESRF. These investigations were supported by the Netherlands Foundation for Chemical Research (CW) with financial aid from the Netherlands Organization for Scientific Research (NWO). We thank the EU for supporting the work at EMBL Hamburg through the HCMP Access to Large Installations Project.

Correspondence and requests for materials should be addressed to B.W.D. Coordinates have been deposited in the Protein Data Bank under accession numbers 1QD5 and 1QD6.

The reaction cycle of isopenicillin N synthase observed by X-ray diffraction

Nicolai I. Burzlaff^{*,†}, Peter J. Rutledge^{*,†}, Ian J. Clifton^{*}, Charles M. H. Hensgens^{*}, Michael Pickford[‡], Robert M. Adlington^{*}, Peter L. Roach^{*} & Jack E. Baldwin^{*}

^{*} The Dyson Perrins Laboratory and the Oxford Centre for Molecular Sciences, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

[‡] The Laboratory of Molecular Biophysics, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

[†] These authors contributed equally to this work

Isopenicillin N synthase (IPNS), a non-haem iron-dependent oxidase, catalyses the biosynthesis of isopenicillin N (IPN), the precursor of all penicillins and cephalosporins¹. The key steps in this reaction are the two iron-dioxygen-mediated ring closures of the tripeptide δ -(L- α -aminoadipoyl)-L-cysteinyl-D-valine (ACV).

It has been proposed that the four-membered β -lactam ring forms initially, associated with a highly oxidized iron(IV)-oxo (ferryl) moiety, which subsequently mediates closure of the five-membered thiazolidine ring². Here we describe observation of the IPNS reaction in crystals by X-ray crystallography. IPNS-Fe²⁺-substrate crystals were grown anaerobically^{3,4}, exposed to high pressures of oxygen to promote reaction and frozen, and their structures were elucidated by X-ray diffraction. Using the natural substrate ACV, this resulted in the IPNS-Fe²⁺-IPN product complex. With the substrate analogue, δ -(L- α -aminoadipoyl)-L-cysteinyl-L-S-methyl-cysteine (ACmC) in the crystal, the reaction cycle was interrupted at the monocyclic stage. These mono- and bicyclic structures support our hypothesis of a two-stage reaction sequence leading to penicillin. Furthermore, the formation of a monocyclic sulphoxide product from ACmC is most simply explained by the interception of a high-valency iron-oxo species.

A challenge in protein crystallography is the possibility of directly observing the structural changes occurring in enzymes during the catalytic cycle. One approach has been the use of Laue crystallography, which has permitted the observation of relatively small changes in structure, such as the photodissociation of a small ligand⁵. The IPNS-catalysed oxidative synthesis of IPN involves significant movements of atoms, through the formation of the two fused heterocyclic rings of penicillin from the linear precursor ACV. Our approach involved the direct oxygenation of anaerobically grown crystals of IPNS complexed with Fe²⁺ and substrates, then structural investigation by X-ray crystallography.

In solution, IPNS shows relatively slow kinetic parameters ($k_{\text{cat}} = 4.6 \text{ s}^{-1}$, $K_m = 0.56 \text{ mM}$ for ACV, where k_{cat} is the maximum catalytic rate when substrate is saturating, and K_m is the concentration of substrate at which the reaction rate is half its maximum value⁶, and initial experiments with crystals exposed to atmospheric pressures of oxygen (air) proved unsuccessful. Therefore, a 'bomb' suitable for exposing crystals to oxygen at pressures up to 60 bar was built^{7,8}, and used to expose crystals of IPNS-Fe²⁺-ACV to 40-bar oxygen for different time periods. After an exposure time of 20 min, a transient pale yellow colour ($\lambda_{\text{max}} 410 \text{ nm}$) was observed uniformly across the crystals; the colour faded with prolonged exposure and was no longer discernible after 320 min. The origin of this chromophore is at present uncertain, but a similar absorption ($\lambda_{\text{max}} 430 \text{ nm}$) has been attributed to a ferryl species in methane mono-oxygenase⁹, although that enzyme has a di-iron centre.

A single crystal of IPNS-Fe²⁺-ACV exposed to the same oxygen pressure for 320 min allowed determination of the structure of IPNS-Fe²⁺-IPN to 1.35 Å resolution (Fig. 1a, b). The structure is primarily that of an IPNS-Fe²⁺-IPN product complex, but residual electron density at the carbonyl and sulphide regions of the ACV cysteine unit make it apparent that the crystal still contains 30% of unreacted IPNS-Fe²⁺-ACV (Fig. 1b). Attempts to obtain complete turnover by extended pressurization under oxygen (for up to 640 min) were unsuccessful. Crystals exposed for these longer times showed low occupancy in the substrate binding region, consistent with slow release of the product from the active site into the surrounding solvent.

The most significant differences between the IPNS-Fe²⁺-IPN structure and the IPNS-Fe²⁺-ACV complex are in the positions of the β -lactam carbonyl group and the cysteinyl sulphur atom. The carbonyl group has moved 60° around the axis of the cysteinyl C₁–C₂ bond, towards Phe 211. The absence of well-ordered hydrogen bonds to this carbonyl group in the substrate complex appears necessary to facilitate this movement. During the reaction, the sulphur migrates around the iron towards the valine β -carbon and out of the site *trans* to His 270, forming the pentacoordinate product complex. A cavity apparent in the IPNS-Fe²⁺-ACV complex accommodates the sulphur in the product structure. The valine isopropyl group has rotated about the C _{α} –C _{β} bond to facilitate formation of the carbon–sulphur bond, and this rotation reverses

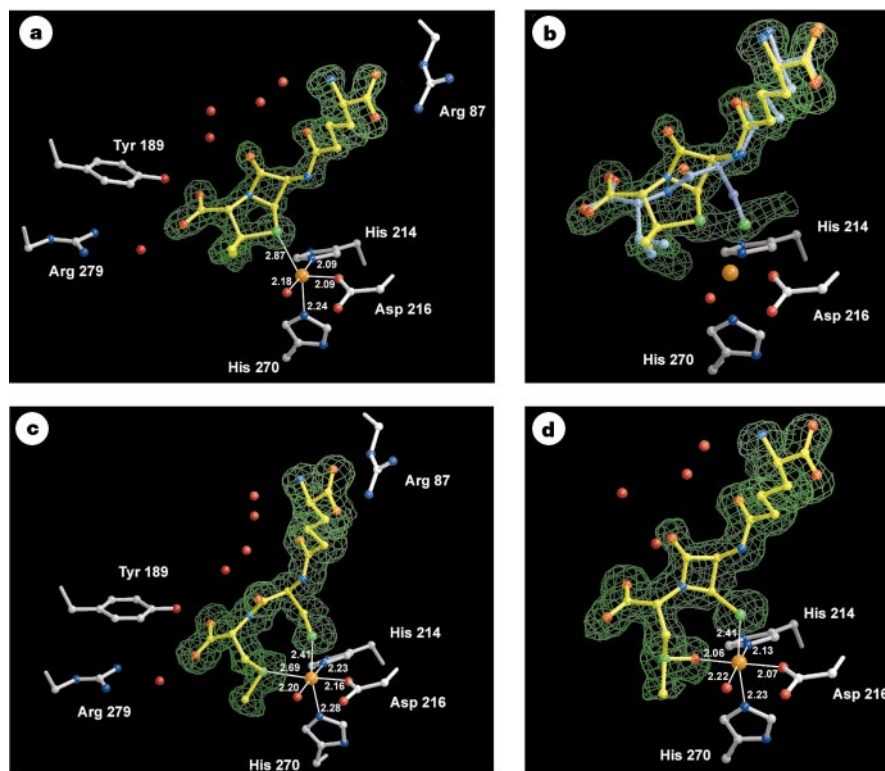


Figure 1 Structures showing changes in the active-site region. All electron-density maps are contoured at 1σ ; distances are shown in Å. **a**, The IPN model with an electron-density map ($2m(10/7(F_o - 0.3F_{\text{calc}})) - DF_c$) of the IPN portion (70%). **b**, The mixed IPN and ACV model with a ($2mF_o - DF_c$) electron-density map. **c**, The IPNS-Fe²⁺-ACmC structure

before exposure to high-pressure oxygen, with a ($2mF_o - DF_c$) electron-density map. **d**, The IPNS-Fe²⁺-ACmC structure after 10-min exposure to oxygen, showing the monocyclic sulphoxide product and a ($2mF_o - DF_c$) electron-density map. *m*, figure of merit; *D* approximates to the fraction of the calculated structure factor that is correct.

the position of the methyl groups bound to it, as thiazolidine formation is known to proceed with retention¹⁰ (Fig. 1a, b).

The IPNS-Fe²⁺-IPN structure also provides some evidence concerning product release from the active site. The IPN carboxylate groups derived from the amino adipoyl side chain and valine remain tethered to the protein. Thus the penicillin, shortened relative to ACV, accommodates less well the three-point attachment to Arg 87, Tyr 189 and iron, and is presumably therefore 'activated' towards dissociation from the active site.

In parallel with our attempts to generate and characterize the bicyclic product in the active site of IPNS crystals, we designed a modified substrate that could, in principle, interrupt the reaction cycle at the monocyclic stage. There is much indirect evidence to support a mechanism involving a monocyclic β -lactam intermediate^{2,11}, possibly coupled to an iron(IV)-oxo species (Fig. 2, 2) en route to the enzyme-product complex (Fig. 2, 3). Given this, and the ease with which IPNS accommodates changes in the valine moiety of ACV¹¹, we selected the modified peptide ACmC as a substrate analogue potentially able to interrupt the reaction cycle as required. The only structural change from ACV is replacement of D-valine (Fig. 2, 1) by D-S-methyl-cysteine (Fig. 2, 4), and our plan was to intercept the proposed iron(IV)-oxo species (Fig. 2, 5) by oxygen transfer to the methyl sulphide group, generating a sulphoxide product (Fig. 2, 6). This is a known reaction of sulphides with metal-oxo species¹².

The modified peptide ACmC was crystallized anaerobically with IPNS and Fe²⁺, and the structure of IPNS-Fe²⁺-ACmC was determined to a resolution of 1.50 Å. The substrate analogue binds in a similar manner to ACV (Figs 1c and 3), being tethered by a salt bridge to Arg 87, by hydrogen bonding to Tyr 189, Arg 279 and Ser 281, and to the iron atom through its cysteinyl sulphur. However, whereas the iron is pentacoordinate in the IPNS-Fe²⁺-ACV complex, in this structure the metal is hexacoordinate: the sulphur

of the S-methyl moiety is bound in the proposed oxygen binding site, *trans* to Asp 216 (ref. 4). The iron-sulphur bond length is 2.69 Å for the sulphide, compared to the 2.41 Å observed for the cysteinyl sulphur-iron distance.

Despite this ligation of the sulphide in the putative oxygen binding site, the application of high-pressure oxygen brought about turnover in the active site of the IPNS-Fe²⁺-ACmC crystals. The crystals were exposed to 20-bar oxygen for a wide range of time periods, and ten minutes proved optimal. The resulting X-ray structure, processed to a resolution of 1.45 Å, clearly shows essentially complete conversion to a monocyclic β -lactam species in the active site of IPNS (Fig. 1d).

We note that the carbonyl oxygen of the β -lactam has moved to a position similar to that seen in the IPNS-Fe²⁺-IPN structure, whereas the cysteinyl sulphur atom occupies an almost identical position to that seen in the starting material. Changes in the region of the S-methylcysteinyl side chain are consistent with the formation of a metal-bound sulphoxide product. When the monocyclic β -lactam sulphoxide (Fig. 2, 6) was fitted to this density, the result was a structure with an oxygen-sulphur distance of 1.47 Å and an iron-oxygen distance of 2.06 Å (Fig. 1d). This structure of exposed IPNS-Fe²⁺-ACmC provides direct evidence for the initial formation of a monocyclic β -lactam species in the reaction cycle of IPNS. Additionally, although we have not directly observed the iron(IV)-oxo species, observation of the sulphoxide directly bound to iron in the site *trans* to Asp 216, the postulated oxygen binding site⁴, is in accord with the intermediacy of a ferryl oxygen species.

It is significant that in both the bicyclic and monocyclic structures the carbonyl oxygen atom of the β -lactam ring is hydrogen-bonded to two well ordered water molecules. In contrast, the corresponding carbonyl oxygen atom in the substrate structures has no hydrogen-bonding interactions with the enzyme or with well ordered water molecules. Such hydrogen bonds to the cyclized structures would

Table 1 Data collection and statistics

	IPNS.Fe ²⁺ .IPN		IPNS.Fe ²⁺ .ACmC		IPNS.Fe ²⁺ .ACmC	
Exposure to oxygen	320 min, 40 bar		Unexposed		10 min, 20 bar	
X-ray source	BW7B, EMBL, Hamburg		9.6, SRS, Daresbury		BW7B, EMBL, Hamburg	
Wavelength λ (Å)	0.8345		0.8700		0.8382	
R_{cryst} (%) [*]	13.1 (8–1.35 Å)		17.3 (25–1.50 Å)		13.3 (8–1.40 Å)	
R_{cryst} (%) [*] ($\ a\ > 2$)	10.5		n.d.		10.5	
R_{free} (%) [†]	18.7		19.9		20.5	
R_{free} (%) [†] ($\ a\ > 2$)	15.3		n.d.		16.9	
B factors [‡]	14.01, 17.61, 17.55		11.83, 13.65, 9.85		14.50, 18.86, 21.51	
Residues	327 (5 to 331)		328 (4 to 331)		328 (4 to 331)	
Water molecules [§]	490 (30)		413 (86)		413 (50)	
Rounds REFMAC ¹⁵	12		21		17	
Rounds SHELXL ¹⁶	10		0		9	
Unit cell (Å)	$a = 46.67, b = 70.91, c = 100.73$		$a = 46.81, b = 71.66, c = 101.63$		$a = 46.78, b = 71.19, c = 101.25$	
Resolution shell (Å)	1.37–1.35		1.58–1.50		1.48–1.40	
Measurements	9,310		5,146		6,601	
Average $\ a\ $	1.8		3.4		2.2	
Unique reflections	1,595		3,572		4,482	
Completeness (%)	43.0		57.3		50.6	
$\ a\ > 3$ data only (%)	6.0		12.6		3.4	
R_{merge} (%) [¶]	29.0		20.0		34.3	

n.d., not determined.

^{*} $R_{\text{cryst}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}| \times 100$

[†] R_{free} = based on 5% of the total reflections.

[‡]Average B factors in order: main-chain, side-chain and substrate.

[§]Total number of water molecules (followed by the number of those half occupied).

[¶]Estimated value.

$R_{\text{merge}} = \sum_i \sum_j |I_{ij} - \langle I_i \rangle| / \sum_i \sum_j I_{ij} \times 100$

necessarily reduce the electron density on the β -lactam nitrogen atom, and thereby stabilize the C–S bond against fragmentation (Fig. 4). This type of fragmentation has been found in penicillins and monocyclic lactams in reactions that cause positive polarization of the sulphur atom^{13,14}. On the other hand, such reduction of electron density on the nitrogen in ACV would inhibit the reaction that forms the β -lactam C–N bond. Thus the timing of these hydrogen-bonding interactions, and the positioning of these water molecules, may well have a crucial role in the overall reaction.

The introduction of dioxygen as a high-pressure gas, coupled to the use of anaerobically grown crystals containing modified substrates, provides an approach to the structural exploration of the reaction cycles of these oxidative enzymes. This methodology allows the acquisition of a series of three-dimensional electron density ‘snapshots’ at selected points along the reaction pathway. In this way we have obtained structures of the enzyme–substrate complex⁴, the complex of enzyme, substrate and NO as an analogue of dioxygen⁴, and now the enzyme–bicyclic product complex, and—using a modified substrate—an enzyme–monocyclic product complex. □

Methods

Oxygenation experiments

A device suitable for exposing crystals to oxygen at pressures up to 60 bar was designed and built (N.I.B. *et al.*, manuscript in preparation). Crystals of IPNS.Fe²⁺.ACV and IPNS.Fe²⁺.ACmC were grown under anaerobic conditions as reported previously³. Exposed crystals were prepared by rapidly transferring a single glass coverslip bearing a drop of solution containing crystals to the pressure vessel of the ‘bomb’. The crystals were reacted by diffusion with 20- and 40-bar oxygen for a range of reaction times (5–640 min), then removed from the ‘bomb’ and rapidly frozen. Safety precautions were taken to minimize the risk of fire and explosions with high-pressure oxygen. Ultraviolet spectra of single crystals were recorded as described¹⁵.

Data collection

Data were collected at 100 K using synchrotron radiation and MAR Research detectors at DESY/EMBL beamline BW7B and SRS Daresbury beamline 9.6 (Table 1). Crystals were approximately 0.5 mm $\times$ 0.2 mm $\times$ 0.1 mm in size. Data were collected in dose mode, taking $\sim$ 30 s per exposure.

Structure determination

For all structures, data were processed with DENZO and SCALEPACK¹⁶, and refined with the CCP4 suite of programs¹⁷. For the two exposed structures, final refinement was

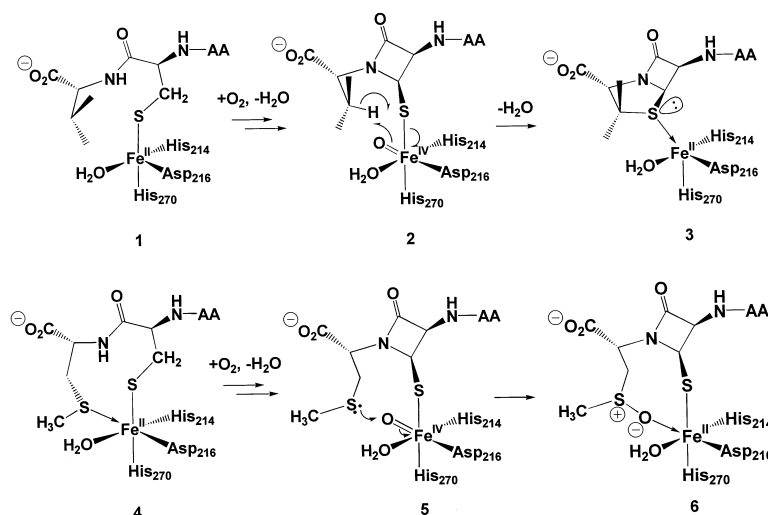


Figure 2 Proposed mechanisms for the oxidation of ACV and ACmC to bicyclic and monocyclic products, respectively. See text for details of compounds 1–6. AA, L- δ -(α -aminoadipoyl).

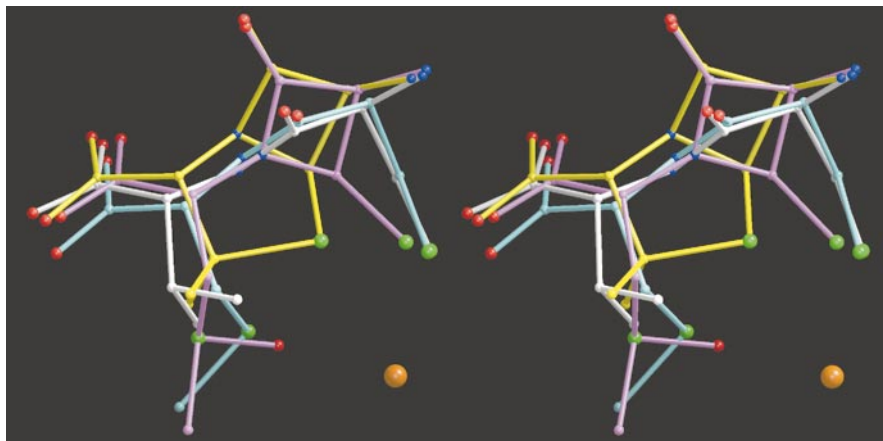
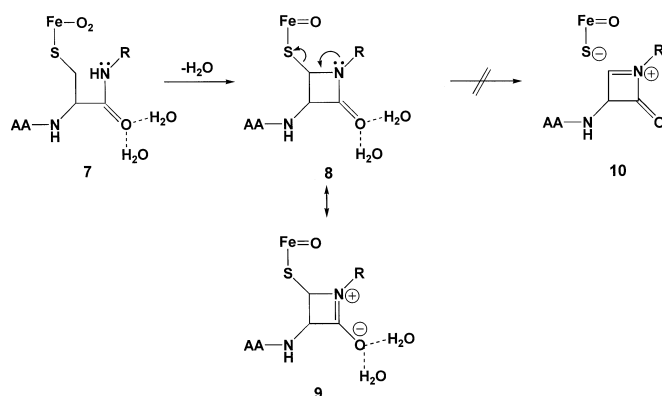


Figure 3 Stereo views of the two substrates and two products overlaid. The key regions that participate in the reaction and the iron atom (orange) are shown; the aminoacylpol side chain, which does not move significantly, is omitted for clarity. Shown are ACV



undertaken using SHELXL98 (ref. 18). Initial structures were obtained by rigid-body refinement of the IPNS-Fe²⁺-ACV main-chain residues into the new unit cell. Electron-density maps were interpreted using program O (ref. 19). The iron–ligand bond lengths were unrestrained throughout refinement, and there were no Ramachandran outliers. Restraints for IPN and the monocyclic sulphoxide were calculated from models obtained from the Cambridge Structural Database. For the two exposed structures, all non-hydrogen atoms were refined anisotropically.

Water molecules, the iron atom and sulphate ions were added during the course of refinement. For each of the three structures, the IPNS-Fe²⁺-ACV main-chain residues were used as the starting model for refinement. In all cases, electron density for the substrate-derived species in the active site was clearly visible throughout refinement. For the IPNS-Fe²⁺-IPN structure, IPN was modelled in during the initial REFMAC¹⁷ refinement, but for the subsequent refinements with SHELXL98 (ref. 18), a model with an IPN:ACV occupancy ratio of 70:30 was used. For the unexposed IPNS-Fe²⁺-ACmC structure, the tripeptide was modelled in after the fifth cycle of refinement. For the exposed IPNS-Fe²⁺-ACmC structure, initial refinement fitting the unreacted substrate analogue to the active-site region showed that reaction had proceeded close to completion. Thus the monocyclic species was modelled in 100% occupancy after the ninth cycle of refinement.

Received 30 April; accepted 16 August 1999.

- Baldwin, J. E. & Abraham, E. Biosynthesis of penicillins and cephalosporins. *Nat. Prod. Rep.* **5**, 129–145 (1988).
- Baldwin, J. E., Adlington, R. M., Moroney, S. E., Field, L. D. & Ting, H. H. Stepwise ring closure in penicillin biosynthesis. Initial beta-lactam formation. *J. Chem. Soc. Chem. Commun.* 984–986 (1984).
- Roach, P. L. *et al.* Anaerobic crystallisation of an isopenicillin N synthase. Fe(II). substrate complex demonstrated by X-ray studies. *Eur. J. Biochem.* **242**, 736–740 (1996).
- Roach, P. L. *et al.* Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827–830 (1997).
- Hajdu, J. & Andersson, I. Fast crystallography and time-resolved structures. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 467–498 (1993).
- Kriauciunas, A. *et al.* The functional role of cysteines in isopenicillin N synthase. *J. Biol. Chem.* **266**, 11779–11788 (1991).
- Berendzen, J. *et al.* Crystal structures of reaction intermediates in cytochrome P450. *Biophys. J.* **74**, A250 (1998).
- Schlichting, I. *et al.* Crystal structures of intermediates occurring along the reaction pathway of cytochrome P450(cam). *FASEB J.* **11**, P2 (1997).
- Lee, S.-K., Nesheim, J. C. & Lipscomb, J. D. Transient intermediates of the methane monooxygenase catalytic cycle. *J. Biol. Chem.* **268**, 21569–21577 (1993).

(white), IPN (yellow), ACmC (blue) and its monocyclic sulphoxide product (pink). Figures were prepared using the programs MOLSCRIPT²⁰ and Raster3D (ref. 21).

Figure 4 The beta-lactam ring is stabilized against fragmentation by hydrogen bonds from well-ordered water molecules observed in both the product structures. Such water are molecules absent from the unexposed IPNS-Fe²⁺-ACV and IPNS-Fe²⁺-ACmC complexes. Movement of the cysteinyl carbonyl group upon oxygen binding (7) and beta-lactam formation enable hydrogen bonding of the carbonyl oxygen to two water molecules (8). This would reduce the electron density on the beta-lactam nitrogen atom (9) and so stabilize the C–S bond against fragmentation (10). R is the remaining portion of either the D-valine or D-S-methylcysteine residue.

- Baldwin, J. E., Adlington, R. M., Domayne-Hayman, B. P., Ting, H.-H. & Turner, N. J. Stereospecificity of carbon–sulphur bond formation in penicillin biosynthesis. *J. Chem. Soc. Chem. Commun.* 110–113 (1986).
- Baldwin, J. E. & Bradley, M. Isopenicillin N synthase: mechanistic studies. *Chem. Rev.* **90**, 1079–1088 (1990).
- Rigau, J. J., Bacon, C. C. & Johnson, C. R. The stereochemistry of oxidation at sulfur. Oxidation of 2-methylthiolane. *J. Org. Chem.* **35**, 3655–3657 (1970).
- Baldwin, J. E. *et al.* Identification and characterisation of shunt metabolites from isopenicillin N synthase. *J. Chem. Soc. Chem. Commun.* 1125–1128 (1988).
- Baldwin, J. E. *et al.* gamma-Lactam formation from tripeptides with isopenicillin N synthase. *J. Chem. Soc. Chem. Commun.* 1128–1130 (1988).
- Hadfield, A. & Hajdu, J. A fast and portable microspectrophotometer for protein crystallography. *J. Appl. Crystallogr.* **26**, 839–842 (1993).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Sheldrick, G. M. & Schneider, T. R. SHELXL: High-resolution refinement. *Methods Enzymol.* **277**, 319–343 (1997).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Kraulis, P. J. Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Murphy, M. E. P. Raster3D Version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Acknowledgements

We thank K. Harlos, C. Schofield, A. Long, J. Elkins, J. Ogle, P. Wright, R. Wilmouth, S. Lee, A. Salmeen, J. Pitt, J. Keeping and the scientists at SRS Daresbury, EMBL Hamburg and ESRF Grenoble for help and discussions. Financial support was provided by the MRC, BBSRC and ESRC. P.L.R. thanks the Royal Society for financial support. N.I.B. was supported by a German DAAD fellowship. P.J.R. thanks the Rhodes Trust for support.

Correspondence and requests for materials should be addressed to J.E.B. (e-mail: jack.baldwin@chem.ox.ac.uk.). The crystallographic coordinates and structure factors have been deposited in the Brookhaven Protein Data Bank (accession nos 1qje and 1qjef for IPNS-Fe²⁺-IPN, 1qiq and 1qiqsf for unexposed IPNS-Fe²⁺-ACmC, and 1qjf and 1qjfsf for exposed IPNS-Fe²⁺-ACmC).

Genes amplified, detected, visualized

Clean PCR, automatic karyotyping, and other gene-linked matters.

Lyse-N-Go

Pierce Chemical www.piercenet.com

Clean PCR preparation

PCR preparation can be done quickly and with low risk of contamination with the ready-to-use Lyse-N-Go PCR Reagent. It provides PCR-ready DNA from cells or tissue in 10 minutes or less and allows direct use of lysed samples for amplifying genomic or plasmid DNA in PCR, eliminating the need for DNA purification kits and purification protocols.

Reader Service No. 100

OmniGrid

GeneMachines www.genemachines.com

Protection for microarrays

GeneMachines has developed a hard-sided enclosure option for their OmniGrid microarraying instrument. The enclosure includes a HEPA filter to reduce dust levels during operation, eliminating the problems that dust can cause for microarraying. The new option also allows the user to control humidity levels to prevent evaporation during spotting. The OmniGrid slide platter is accessible through one hinged door and two sliding side doors in the enclosure, allowing

easy access to sample plates and slides. Future upgrades for OmniGrid will be compatible with the new enclosure.

Reader Service No. 101

CryoProbe

Bruker NMR www.bruker.com

Biological NMR for structural genomics and proteomics

The NMR triple-resonance CryoProbe is optimized for high-field biological NMR applications. CryoProbe is a 5-mm 500-MHz inverse probe for ^1H detection, with simultaneous ^{13}C and ^{15}N decoupling. Bruker claim an S/N improvement over conventional probes of a factor of 3–4, and an increase in throughput by a factor of 9–16.

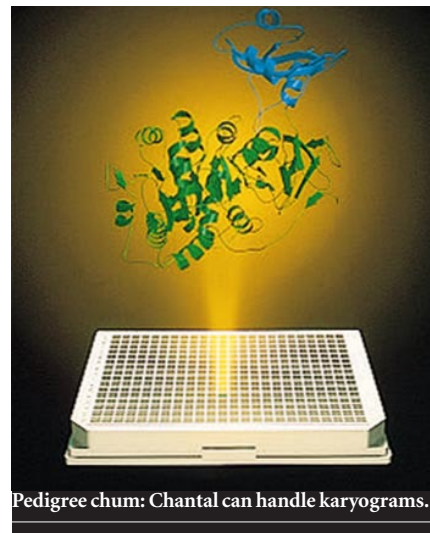
Reader Service No. 102

Image editor

Leica Microsystems
www.leica-microsystems.com

Automatic chromosome classification

Newly included in the Leica Chantal chromosome karyotyping system is the image editor, capable of producing pedigree analysis diagrams and partial karyotypes with both fluo-



Pedigree chum: Chantal can handle karyograms.

rescent and brightfield images. Leica Chantal is compatible with Windows NT and 98 and supports dual-display mode. The Image Editor allows the user to design partial karyograms and pedigree charts.

Reader Service No. 103

TransIT

Mirus Corporation www.genetransfer.com

Designed for in vivo gene delivery

This *in vivo* gene delivery system contains a polymer formulation to facilitate efficient delivery of exogenous DNA complexes into tissues of live mice. TransIT consists of two components: a proprietary polymer formulation that complexes with the DNA to be delivered, and a dilution/injection reagent for the DNA delivery step. The DNA is injected via the tail vein.

Reader Service No. 104

TSA Plus

NEN Life Science www.nenlifesci.com

Signal amplification without PCR

TSA Plus employs novel signal amplification technology to enhance the sensitivity of chro-

LipofectAmine

Life Technologies

www.lifetech.com

A broad spectrum transfection reagent

The LipofectAmine 2000 reagent was developed to provide high gene and protein expression levels for a wide variety of adherent and suspension cell lines, including COS-7, 293, and CHO. The reagent functions in the presence or absence of serum, using a quick and simple protocol. For high-throughput screening, assay times may be further reduced by adding LipofectAmine 2000 Reagent-DNA complexes directly into 96-well plates at the time of cell seeding.

Reader Service No. 112



High activity, broad spectrum.

ADVERTISEMENT



Shaping the future of spectrophotometry

- *lightens workloads*
- *increases throughput*
- *creates more space*

lightwave

UV/Vis Diode Array Spectrophotometer

www.wpaltd.co.uk





READER ENQUIRY NO. 26

mogenic detection for *in situ* hybridization without the need for PCR. It is completely biotin-free and optimized for use with digoxigenin probe labelling systems. TSA Plus also works in immunohistochemistry protocols, providing equivalent amplification benefits. **Reader Service No. 105**

LucLite Plus

Packard www.packardinst.com
A reagent kit for 384-well microplate-based reporter gene assays

The LucLite Plus kit is designed to simplify luciferase reporter gene assays in higher density microplate formats. The kit provides a luminescent signal 250% brighter than is possible with the previous LucLite technology, for easy detection of 8 fg of firefly luciferase.

Reader Service No. 106

DASH

Hybaid www.hybaidd.co.uk
Single nucleotide polymorphism scoring

The DASH (Dynamic Allele Specific Hybridization) system automates SNP detection using a streptavidin-coated microtitre plate, so it is compatible with automation. DASH utilizes fluorescent signal detection based around dynamic heating, and coincident monitoring of DNA denaturation. PC-based software provides automatic scoring

of SNPs following predetermined criteria.

Reader Service No. 107

OmniBank

Lexicon Genetics Inc. www.lexgen.com
Mouse clones on tap

Lexicon Genetics are running a subscription Internet program to enable commercial and academic non-profit institutions to access OmniBank, the company's proprietary genomics library and database. Through the Lexicon Genetics web site, subscribers can reach 50,000 mouse clones, an extensive collection of knockout mouse embryonic stem cells and gene sequence information.

Reader Service No. 108

DNAzol

Molecular Research Center
www.mrcgene.com

Genomic DNA isolation

DNAzol is a ready-to-use reagent for the isolation of genomic DNA from solid and liquid samples of animal and plant origin. The reagent contains a guanidine-detergent lysing mixture that promotes selective precipitation of DNA. Genomic DNA isolation can be completed in 30 minutes with DNA recovery of 70–100%. Fifty ml of DNAzol is adequate to process 50 samples of 25–50 mg of tissue.

Reader Service No. 109

Mutation detection

CBS Scientific www.cbssci.com
Single base changes detected

Five Denaturing Gradient Gel Electrophoresis (DGGE) systems for detecting single base changes and polymorphisms in genomic, cloned, and PCR amplified DNA are now available from CBS. Both single and dual gel cassette units are available, and each can accommodate two to eight gels simultaneously. The systems are suitable for fluorescence based assays when used with the optional borosilicate glass plate sets.

Reader Service No. 110

GeneSpring

Silicon Genetics www.sigenetics.com
A new version of this software package

The GeneSpring analysis and visualisation package for genomic expression analysis is now at Version 2.1. GeneSpring is both micro array and Affymetrix compatible. Features introduced with Version 2.1 include data import via a 'wizard', enhanced overlay and tree building capabilities, and Mac compatibility.

Reader Service No. 111

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

Check out our homepage at <http://www.resgen.com>

Peptides
\$14/residue
10 residue minimum

Polyclonal Antibodies - \$1,195

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 4

Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com

New Probe Sizes
0.45cm

READER ENQUIRY NO. 54

Trichostatin A

CC1=CC=C(C(=C1)C(=O)C(C)C/C=C/C(=O)NO)C

from *Streptomyces*
Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call

Wako BioProducts: (800) 992-9256; (804) 271-7677.

READER ENQUIRY NO. 98

HIGH QUALITY PEPTIDE SYNTHESIS

- 'Standard' and 'non-standard'
- Fast delivery
- Competitive prices
- Also advisory services

Pepsyn Ltd

Contact: Dr J A Smith, Chairman, Pepsyn Ltd,
School of Biological Sciences, Life Science Building,
The University of Liverpool L69 7ZB, UK
Tel: 0151 794 4329 Fax: 0151 794 4349
e-mail: J.A.Smith@liverpool.ac.uk
Website: www.pepsyn.co.uk

READER ENQUIRY NO. 22

CUSTOM PEPTIDES

HPLC MS MALDI-TOF

- Up to 30 aa
- HPLC & MS data

Purity	>80%	>95%	>99%
Amount	15mg	10mg	

info@genosphere-biotech.com
Tel: +33 1 42 71 70 21 • Fax: +33 1 42 71 70 23

GENOSPHERE BIOTECHNOLOGIES

READER ENQUIRY NO. 24

contact: info@TIB-MOLBIOL.de

TIB MOLBIOL

Labelled Oligonucleotides...

...for Real-Time PCR, FISH, fragment analysis and sequencing

TIB MOLBIOL offers a broad spectrum of fluorescent labels, including ABI dyes (FAM, TET, HEX, JOE, TAMRA, ROX), Roche dyes for the LightCycler, Amersham-Pharmacia dyes (Cy3, Cy5, Cy5.5), and all dyes from Molecular Probes (BODIPY, Oregon).

High quality and quick turnaround guaranteed.

READER ENQUIRY NO. 25